

Holt, Stephanie Elspeth (2014) *Target site DNA recognition by Tn3 and Sin resolvases*. PhD thesis.

<https://theses.gla.ac.uk/6215/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>  
[research-enlighten@glasgow.ac.uk](mailto:research-enlighten@glasgow.ac.uk)

**Target site DNA recognition by Tn3 and Sin resolvases**

**Stephanie Elspeth Holt**

**B.Sc. (Hons.)**

**Submitted in fulfilment of the requirements for the Degree of  
Doctor of Philosophy**

**University of Glasgow**

**College of Medical, Veterinary and Life Sciences**

**Institute of Molecular, Cell and Systems Biology**

**2014**

**© S. E. Holt, 2014**

## Summary

This work examined the DNA-recognition and synapsis capabilities of Sin and Tn3 resolvases, by alteration of regions of the protein to determine their significance for resolvase function. The importance of the E-helix (an  $\alpha$ -helical structure in the centre of the resolvase, connecting the catalytic domain and DNA-binding domain of the protein) of Sin recombinase, for DNA recognition at the centre of *res* site I, and for the formation of a synaptic tetramer was investigated. It was observed that swapping of the E-helix 'arm' region from Sin residues to the equivalent residues from Tn3 resolvase abolished formation of the synapse. Truncated variants of Sin with the catalytic domain removed, termed 'tadpoles', were compared with a full-length activated mutant of Sin (Q115R) to determine the importance of the E-helices in formation of the synapse. From the 'tadpole' constructs, it was determined that the E-helix region between residues 109 and 124 is required for the formation of synaptic complexes, as deletion of this region was observed to abolish formation of the complex *in vitro*. Structure-guided single-residue mutations in this region of the E-helix further demonstrated that residues L112, I116 and M119, which are located on the flat hydrophobic interface between the E-helices of the Sin subunits in the synaptic tetramer complex, have a key role in the formation of this complex. Mutation of any of these residues to an alanine, both in the truncated Sin variants with a wild-type background and in the full-length activated mutant Q115R blocks formation of this complex. The truncated versions of Sin support the hypothesis that a bundle of four E-helices comprises a basic synapsis-competent module. In a related project, transcription activator-like effector recombinases (TALERS), chimeric constructs were constructed with the catalytic domain of an activated variant of Tn3 resolvase (NM) fused to the DNA binding domain of a TALE protein, TALERS with different N-terminal TALE truncations were constructed. It was demonstrated that the TALER constructs were successfully targeted to designed cognate 'T-sites', utilising the DNA-binding specificity of the TALE domain. It was also demonstrated that some of the TALERS constructed are able to catalyse cleavage at T-sites *in vitro*.

## **Acknowledgements**

Working in the lab over the last few years to create the data that are to be presented in this thesis was a great learning experience which I have enjoyed. That being said, the actual writing of this thesis was one of the most difficult undertakings of my life. I would like to take this opportunity to thank my awesome family, your patience and understanding during my write-up and support throughout my PhD means so much to me. I also want to thank Marshall for giving me the opportunity to work on this interesting project, and for the suggestions and assistance which you have provided me over the years. And of course a huge thank you to everyone with whom I have worked in the lab – to our brilliant technician Arlene, Femi -I will miss your constant making fun of me in the lab! - Martin, Sally, Sean, Jeff, Jia, Hayley, Steve, Rich and many others. Thank you all for making the last few years such a wonderful experience, for sharing your knowledge and expertise with me and for the awesome times we have shared. Thank you also to the BBSRC for the generous funding of my research.

The research reported in this thesis is my own work unless it is otherwise stated and it has not been submitted for any other degree.

Stephanie E Holt

## **Abbreviations:**

### **Chemicals/Reagents:**

AcOH	Acetic acid
APS	Ammonium persulphate
ATP	Adenosine triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid (disodium salt)
EtBr	Ethidium bromide
KOAc	Potassium acetate
SDS	Sodium dodecyl sulphate
TAE	Tris acetate EDTA (electrophoresis buffer)
TBE	Tris borate EDTA (electrophoresis buffer)
TEMED	N-N-N'-N'-tetramethylethylenediamine
Tris	Tris(hydroxymethyl)aminomethane

### **Commonly used abbreviations:**

DNA	Deoxyribonucleic acid
HR	Homologous recombination
HDR	Homology directed repair
NHEJ	Non-homologous end joining
PAGE	Polyacrylamide gel electrophoresis
RVD	Repeat variable di-residue
TALE	Transcription activator-like effector
TALEN	TALE nuclease
TALER	TALE recombinase
UV	Ultraviolet
ZFN	Zinc finger nuclease
ZFR	Zinc finger recombinase

## Units:

A	Amperes	bp	Basepairs
V	Volts	rpm	Revolutions per minute
W	Watts	OD	Optical density
l	Litres	k	Kilo ( $10^3$ )
m	Metres	c	Centi ( $10^{-2}$ )
g	Grams	m	Milli ( $10^{-3}$ )
h	Hours	$\mu$	Micro ( $10^{-6}$ )
min	Minutes	n	Nano ( $10^{-9}$ )
$^{\circ}\text{C}$	Degrees Celsius	p	Pico ( $10^{-12}$ )
M	Molar		

Summary.....	i
Acknowledgements.....	ii
Abbreviations.....	iv
<b>1. Chapter 1: Introduction.....</b>	<b>1</b>
<b>1.1 Recombination in nature.....</b>	<b>1</b>
<b>1.1.i Homologous recombination.....</b>	<b>1</b>
<b>1.1.ii Site-specific recombination.....</b>	<b>2</b>
<b>1.2 Types of recombinase.....</b>	<b>4</b>
<b>1.2.i Tyrosine recombinases.....</b>	<b>4</b>
<b>1.2.ii Serine recombinases.....</b>	<b>5</b>
<b>1.3 Tn3 resolvase .....</b>	<b>9</b>
<b>1.3.i Recombination by Tn3.....</b>	<b>9</b>
<b>1.3.ii The structure of Tn3/<math>\gamma\delta</math> resolvase.....</b>	<b>13</b>
<b>1.4 Sin resolvase.....</b>	<b>16</b>
<b>1.4.i Recombination by Sin.....</b>	<b>16</b>
<b>1.4.ii The structure of Sin resolvase.....</b>	<b>19</b>
<b>1.5 Genetic applications of site specific recombinases.....</b>	<b>22</b>
<b>1.6 Other methods of genome editing.....</b>	<b>24</b>
<b>1.6.i The CRISPR/Cas System.....</b>	<b>25</b>
<b>1.6.ii Zinc finger nucleases (ZFNs).....</b>	<b>27</b>
<b>1.6.iii TALE nucleases (TALENs).....</b>	<b>28</b>
<b>1.6.iv Zinc finger recombinases.....</b>	<b>33</b>
<b>1.6.v TALE recombinases (TALERs).....</b>	<b>35</b>
<b>1.7 Project aims.....</b>	<b>35</b>
<b>2. Chapter 2: Materials and methods.....</b>	<b>37</b>
<b>2.1 Bacterial strains.....</b>	<b>37</b>
<b>2.2 Bacterial growth media.....</b>	<b>37</b>
<b>2.3 Antibiotics.....</b>	<b>38</b>
<b>2.4 Chemicals.....</b>	<b>38</b>
<b>2.5 Gene synthesis.....</b>	<b>39</b>
<b>2.6 Oligonucleotides.....</b>	<b>39</b>
<b>2.7 Plasmids.....</b>	<b>39</b>
<b>2.7.i High-level resolvase expression plasmids.....</b>	<b>39</b>



2.7.ii Low-level resolvase expression plasmids.....	55
2.7.iii <i>In vitro</i> recombination substrate plasmids.....	55
2.7.iv <i>In vivo</i> recombination substrate plasmids.....	56
2.8 Making <i>E. coli</i> cells competent.....	56
2.8.i Chemically competent cells.....	56
2.8.ii Electro-competent cells.....	58
2.9 Transformation of competent cells.....	58
2.9.i Transformation of chemically competent <i>E.coli</i> cells.....	58
2.9.ii Transformation of electro-competent cells.....	59
2.10 Plasmid DNA preparation.....	59
2.10.i Small scale DNA preparation.....	59
2.10.ii Medium scale DNA preparation.....	60
2.11 Ethanol precipitation of DNA.....	60
2.12 Digestion of DNA by restriction endonucleases.....	60
2.13 Gel electrophoresis.....	61
2.13.i Agarose gel electrophoresis.....	61
2.13.ii Polyacrylamide gel electrophoresis.....	61
2.14 DNA size marker ladder.....	62
2.15 Sample loading buffers.....	62
2.16 Visualisation of DNA by ethidium bromide staining.....	63
2.17 DNA extraction from gels.....	63
2.18 Purification and annealing of oligonucleotides.....	63
2.19 Ligation of DNA fragments from restriction digests.....	64
2.20 Sequencing plasmid DNA.....	64
2.21 <i>In vivo</i> (MacConkey) recombination assay.....	64
2.22 Purification of proteins.....	65
2.23 Laemmli SDS-PAGE.....	68
2.24 <i>In vitro</i> cleavage assays.....	68
2.25 Radioactive labelling of DNA for binding assays.....	69
2.26 Gel purification of labelled DNA.....	70
2.27 Crush and soak.....	70
2.28 Phenol chloroform extraction.....	70
2.29 <i>In vitro</i> Binding assays.....	71

<b>2.30</b>	Crystal structure models.....	71
<b>3.</b>	Chapter 3: Roles of the E-helix in the DNA-binding and synapsis of Sin resolvase.....	72
<b>3.1</b>	Introduction.....	72
<b>3.2</b>	Sin ‘tadpole’ design and construction.....	77
<b>3.3</b>	Binding and synapsis of <i>res</i> site I.....	81
<b>3.4</b>	Sequence selectivity of <i>res</i> site I.....	86
<b>3.5</b>	Sequence selectivity by the E-helix ‘arm’ of resolvase.....	92
<b>3.6:</b>	Mutations of the Sin E-helix.....	101
<b>3.6.i</b>	<i>In vitro</i> binding of E-helix mutants.....	104
<b>3.6.ii</b>	<i>In vivo</i> recombination by E-helix mutants of Sin.....	109
<b>3.7:</b>	Discussion.....	111
<b>4.</b>	Chapter 4: TALE Recombinases.....	114
<b>4.1</b>	Introduction.....	114
<b>4.1.i</b>	TALE proteins.....	114
<b>4.2</b>	Designing new TALERs.....	119
<b>4.3</b>	Construction of TALERs.....	125
<b>4.4</b>	TALER Activity <i>in vivo</i> and <i>in vitro</i> .....	130
<b>4.5</b>	Discussion.....	138
<b>5.</b>	Chapter 5: General discussion and conclusions.....	141
<b>6.</b>	References.....	144
<b>7.</b>	Supplementary data.....	153

# **Chapter 1: Introduction**

## **1.1 Recombination in Nature**

Recombination is the process by which two molecules of DNA (or RNA) are broken and re-joined. Recombination can occur between molecules of DNA which are similar, as in homologous recombination (HR), or different as seen in non-homologous end joining (NHEJ). Bacteria and eukaryotic organisms use recombination as a DNA repair mechanism, for resolution of plasmid multimers and for co-integrate resolution. In eukaryotic organisms, recombination also has a role in meiosis, where chromosomal crossover events lead to increased genetic diversity.

### **1.1.i Homologous recombination**

Homologous recombination involves the exchange of DNA between two regions which share the same or very similar sequence, and requires complex protein machinery. Homologous recombination involves the formation of Holliday junction intermediates, branch migration of the junction and resolution to form recombinant molecules (Holliday 1964; Ariyoshi et al. 2000). This type of recombination is important for maintaining the integrity of the genome, creating new diversity, rescue of stalled replication forks, directing gene expression, controlling the segregation of chromosomes and telomere maintenance (Camerini-Otero and Hsieh; 1995, San Fillipo et al. 2008).

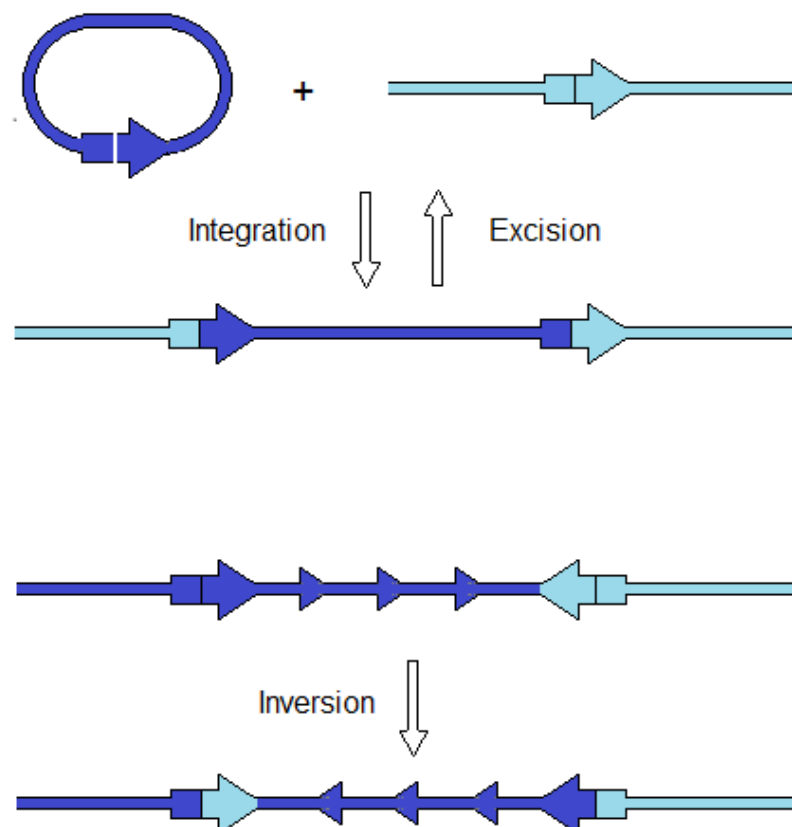
### **1.1.ii Site-specific recombination**

Site-specific recombination involves the reciprocal exchange of DNA strands with limited sequence homology by the breakage, re-arrangement and joining of molecules of DNA. The basic requirements for carrying out a site-specific recombination reaction are two DNA partners containing short recognition sequences (sites) usually between 30 and 200 base pairs in length, and a specialised recombinase protein. Recombinase recognises and binds to its cognate sites, then carries out strand cleavage, exchange and re-ligation of the DNA to a new partner. This reaction stores energy by the formation of a covalent protein-DNA intermediate at strand cleavage, and thus does not require a high energy co-factor such as ATP (Grindley 1994; Grindley et al. 2006).

Site-specific recombinases have the potential to undergo three reaction types; Excision, inversion and integration (Figure 1.1); which of these reactions is carried out is determined and tightly regulated by topological selectivity and site orientation (Grindley et al. 2006). Recombination between two linear DNA molecules yields two linear recombinant products.

The excision and inversion reactions occur between sites within the same molecule of DNA (intramolecular recombination). The excision (resolution) reaction is carried out between two sites which are in direct (head to tail) orientation. Inversion occurs when the sites are instead present in indirect (head to head) repeat as shown in Figure 1.1.

Integration occurs if recombination takes place between different molecules of DNA (intermolecular recombination), as long as one or both of the DNA molecules are circular (Stark et al. 1992).



**Figure 1.1: Potential outcomes of recombination**

Integration results from recombination between parallel sites on separate DNA molecules. Excision results from intramolecular recombination between sites in a head-to-tail (direct) orientation and inversion from recombination between head-to-head (inverted) sites.

Site-specific recombinase enzymes carry out many functions in biological systems, including: integration and excision of DNA (e.g. phage genomes), transfer of gene cassettes, monomerisation of plasmid multimers and resolution of transposition intermediates (Grindley 1994; Grindley et al. 2006). These properties make them excellent targets for study as potential tools for genetic engineering.

## **1.2 Types of recombinase**

There are two main families of recombinase; tyrosine and serine recombinases, named according to the amino acid (aa) residue in their catalytic domain which acts as a nucleophile, forming a covalent protein-DNA linkage during the recombination reaction, which shall be discussed below. Both of these recombinase families use their own distinct mechanism of recombination (Grindley et al. 2006).

The site-specific recombination reaction involves recognition of and binding to specific DNA sites by a recombinase protein, followed by synapsis, cleavage of DNA and displacement of a hydroxyl group on the DNA backbone by a nucleophile on a side chain of the recombinase. The mechanisms for tyrosine and serine recombinases differ.

### **1.2.i Tyrosine recombinases**

Tyrosine family recombinases, including Cre (from the *Escherichia coli* bacteriophage P1) and Flp (from *Saccharomyces cerevisiae*), use an active site tyrosine to create a transient covalent phospho-tyrosine linkage to a scissile phosphate on the DNA backbone. Tyrosine recombinases make a single stranded cut on both molecules of DNA, creating a 3' phospho-tyrosine bond and leaving a free 5' OH. Each free 5' OH then attacks the phospho-

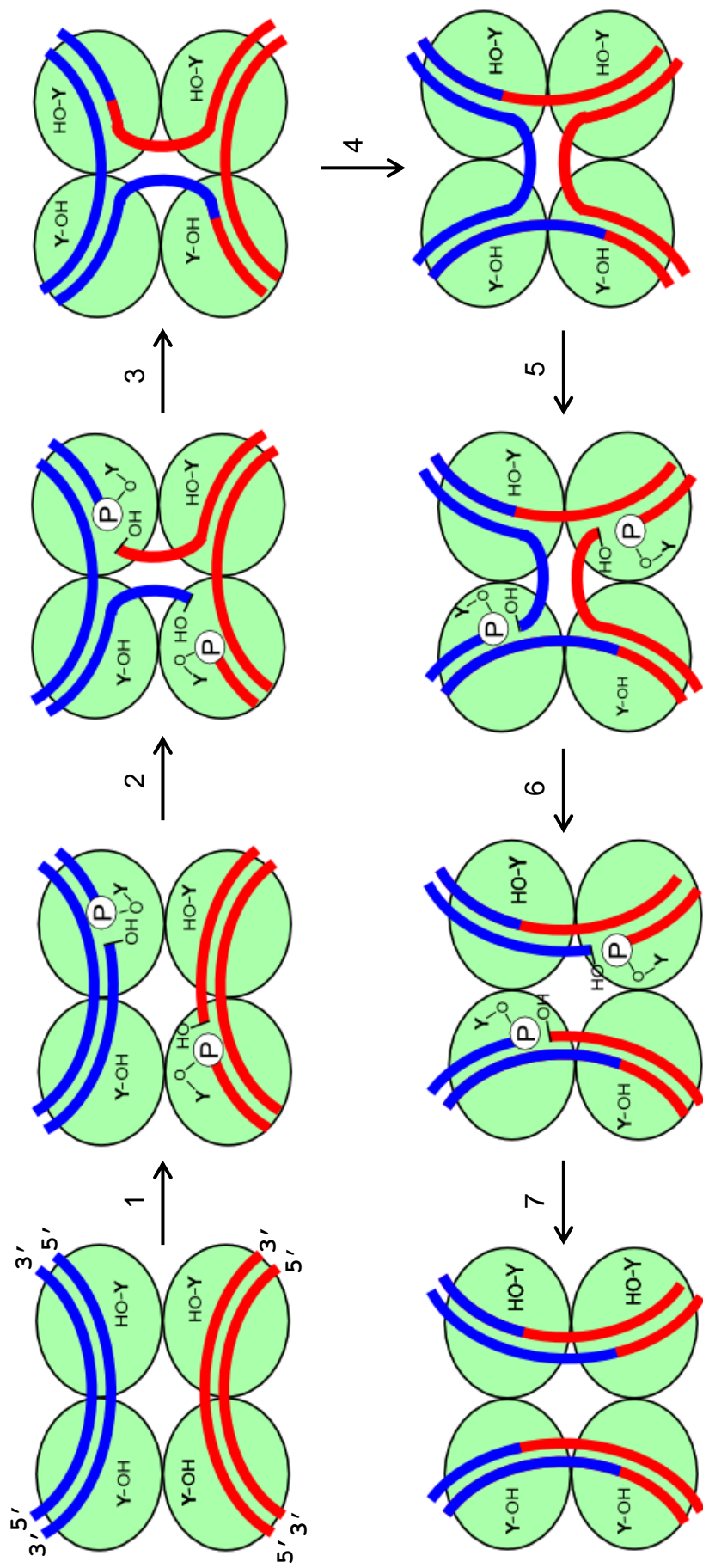
tyrosine linkage on the opposite DNA molecule and a Holliday junction intermediate is formed. The same process is then repeated on the other pair of strands. Following this second round of strand exchange, two recombinant molecules are formed and the recombinase subunits dissociate as shown in figure 1.2 (Grindley et al. 2006).

### **1.2.ii Serine recombinases**

Serine recombinases make a double strand break at the crossover site in each parental DNA duplex; the active site serines of two subunits bound to each site attack the DNA, forming a transient double-strand break, with each 5' end linked to the recombinase via a phospho-serine linkage. Following strand exchange the strands are re-ligated and a new recombinant molecule is formed, as illustrated in figure 1.3 (Stark and Boocock 1995; Grindley 2002; Grindley et al. 2006).

In many systems, including those of Tn3 and Sin to be studied here and the related  $\gamma\delta$  resolvase, the site-specific recombination reaction is tightly regulated to ensure that the appropriate forward reaction is carried out (Grindley et al. 2002). This control is in the form of regulatory subunits which are required to form the synapse necessary for recombination to occur in the WT system. These regulatory subunits bind to the accessory site DNA to regulate the synapsis of the WT recombinase.

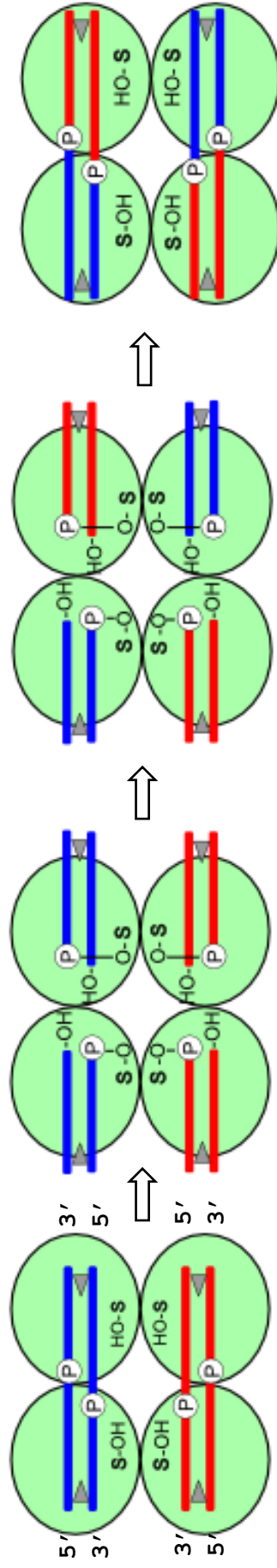
'Activated' mutants of resolvase have been isolated, which promote recombination at site I in the absence of the accessory-sites which are usually required as will be discussed later in this chapter, (Arnold et al 1999, Burke et al 2004, Rowland et al 2009). Such mutants are invaluable tools for the study of protein-DNA interactions at site I. The research carried out for this study utilises two such mutants, Sin Q115R (single mutant) and Tn3 NM (multiple mutant; R2A, E56K, G101S, D102Y, M103I, Q105L) (Rowland et al. 2009; Burke et al. 2004; Olorunniji et al. 2008).



**Figure 1.2: Strand exchange by tyrosine recombinases**

Cartoon depicting strand exchange by tyrosine recombinases. One strand of each head-to-tail DNA duplex is cleaved by the catalytic tyrosine nucleophile (1), forming a transient phospho- tyrosine bond at the 3' end and leaving a free 5' OH, strands are exchanged (2), and a Holliday junction intermediate is formed by ligation of the strands (3). Following isomerisation of this junction (4), the second DNA strands are likewise exchanged and ligated , giving recombinant DNA duplexes (5-7).

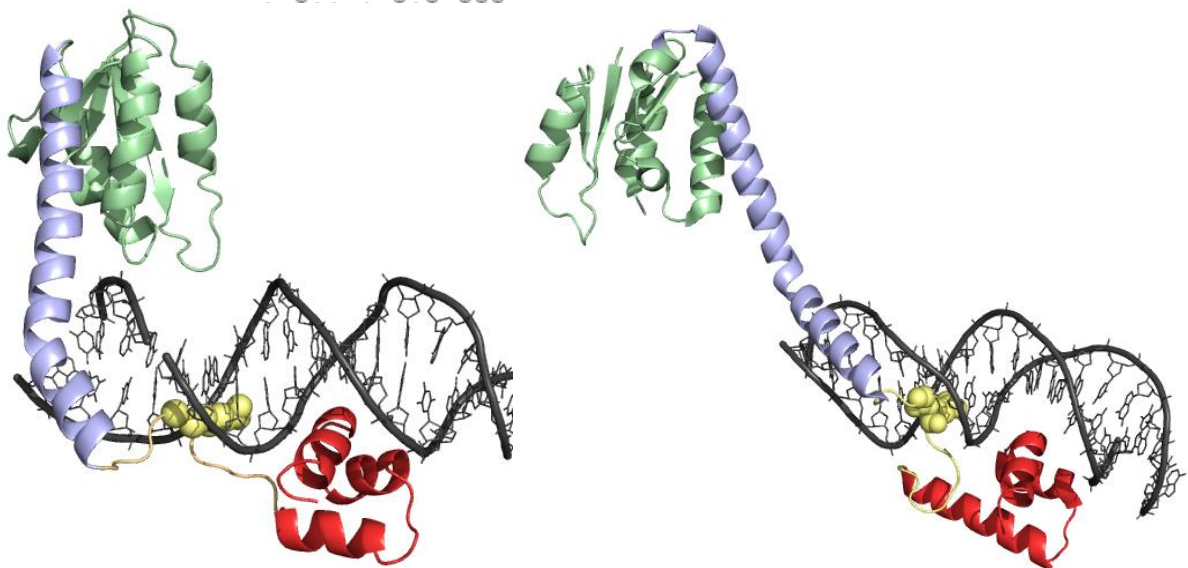
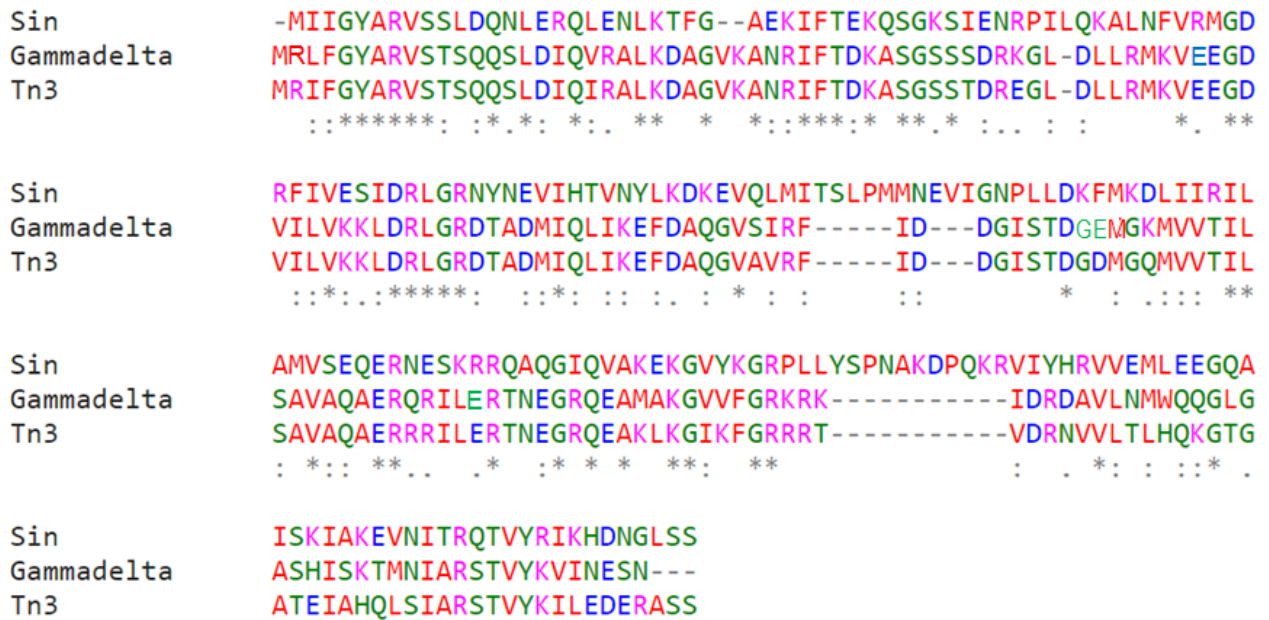




**Figure 1.3: Strand exchange by serine recombinases**

Cartoon depicting strand exchange by serine recombinases: Double-stranded cleavage at each head-to-tail site is carried out by displacement of a scissile phosphate in the DNA backbone by the nucleophiles of two recombinase subunits, forming a transient phospho-serine bond to the DNA 5' end. Following exchange of the DNA half-sites relative to each other, the strands are re-ligated. Re-ligation of the recombinant DNA is the reverse of the cleavage reaction. Inverted repeats are indicated by grey triangles.

The resolvases discussed in this section are related members of the Serine recombinase family.  $\gamma\delta$  resolvase and Tn3 resolvase are closely related, with 78.5% sequence identity, Sin is a more distant relative, with 32% identity to Tn3 and 31% identity to  $\gamma\delta$  (see figure 1.4).



**Figure 1.4: Alignment of  $\gamma\delta$ , Tn3 and Sin**

(A) Sequence alignment showing identity of the resolvases discussed in this work. Tn3 and  $\gamma\delta$  have 78.5% sequence identity, Tn3 and Sin have 32% identity and  $\gamma\delta$  and Sin have 31% identity. Alignment was performed using Clustal Omega multiple sequence alignment software, EMBL-EBI. (B) Structural comparison of Sin(right) and  $\gamma\delta$  (left) resolvases, made using crystal structures 2R0Q and 1GDT respectively.

### 1.3 Tn3 resolvase

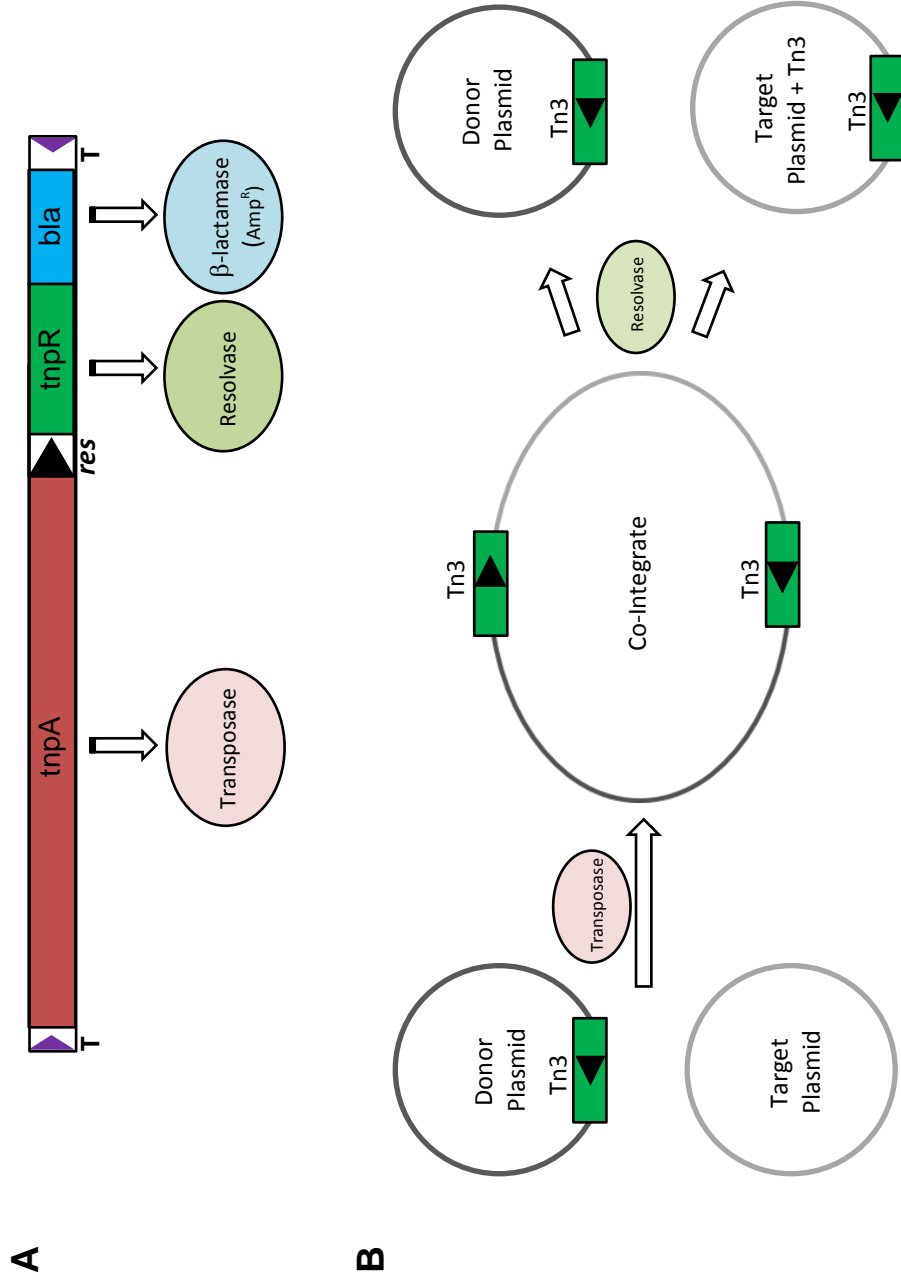
Tn3 resolvase is a member of the serine recombinase family, encoded by the bacterial transposon Tn3, which also encodes  $\beta$ -lactamase (*bla*) and transposase (*tnpA*) genes (figure 1.5). The biological function of Tn3 resolvase in nature is the resolution of co-integrates formed during replicative transposition of its transposon (Grindley et al. 2002).

Each subunit of Tn3 and the related  $\gamma$  resolvase has a structure consisting of a catalytic N-terminal domain, E-helix, AT hook DNA binding motif and C-Terminal DNA-binding domain (figure 1.6) (Yang and Steitz 1995). The structure of resolvase and the importance of the structure for the protein's function will be discussed further in section 1.3.ii

#### 1.3.i Recombination by Tn3 resolvase

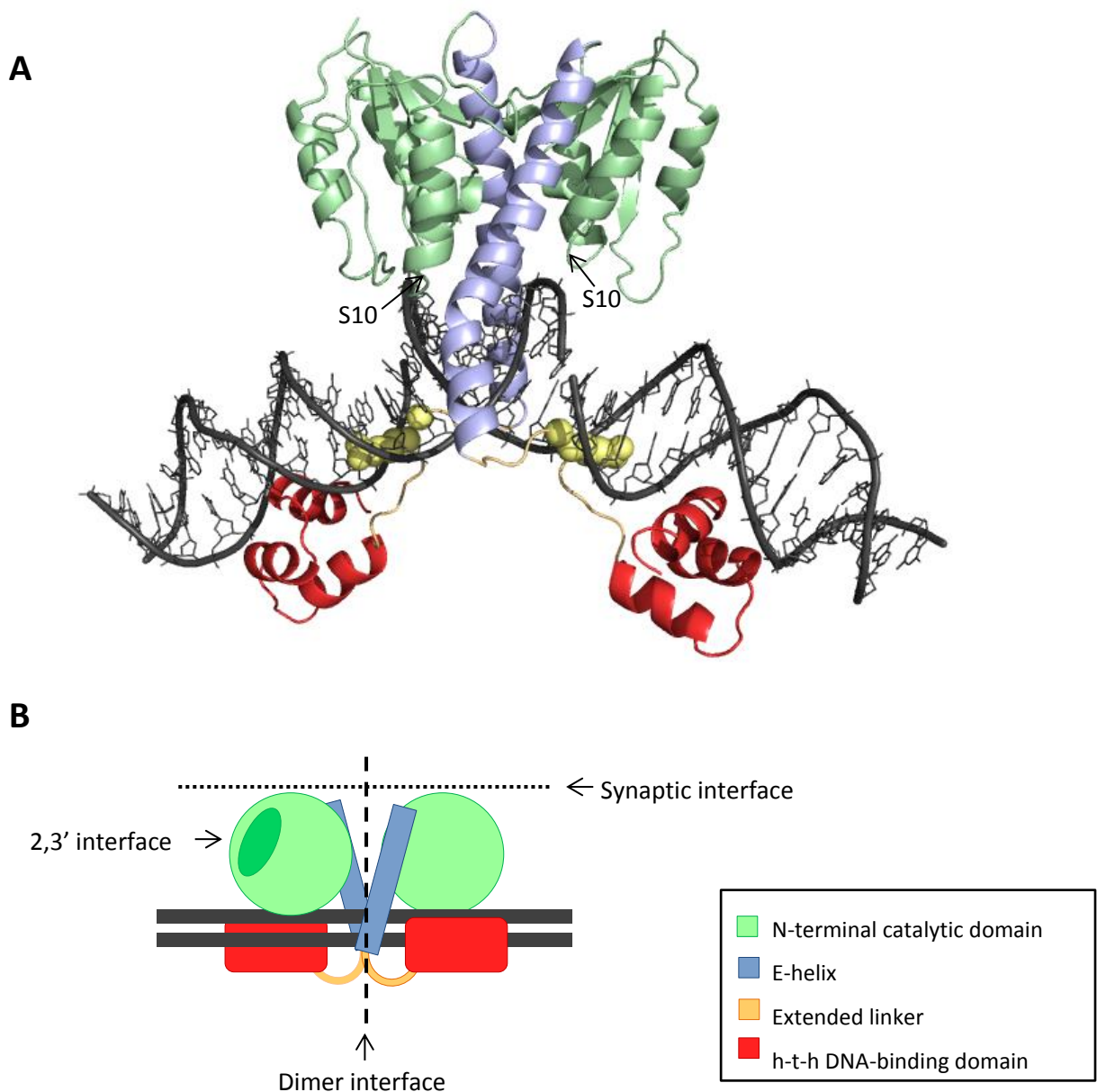
Resolvase-mediated recombination requires an identical pair of DNA sites, known as '*res*'. The full Tn3 *res* site is 114 bp in length, consisting of a 28 bp catalytic site (site I) which has a 12 bp head-to-head repeat sequence at each end and two regulatory sites, site II and site III, of 34 bp and 25 bp in length respectively. A 22 bp spacer separates sites I and II, and a 5 bp spacer separates sites II and III (see figure 1.7) (Burke et al. 2004). Two catalytic subunits of Tn3 resolvase bind to each of the two site I's as dimers, with four more resolvase subunits (two dimers) binding to sites II and III; thus the fully assembled synapse consists of two full *res* sites bound by a total of twelve subunits of resolvase.

Resolution of a circular plasmid by Tn3 resolvase yields a 2-node catenane product as seen in figure 1.7. Regulatory dimers bound at sites II and III in the



**Figure 1.5: The Tn3 transposon:**

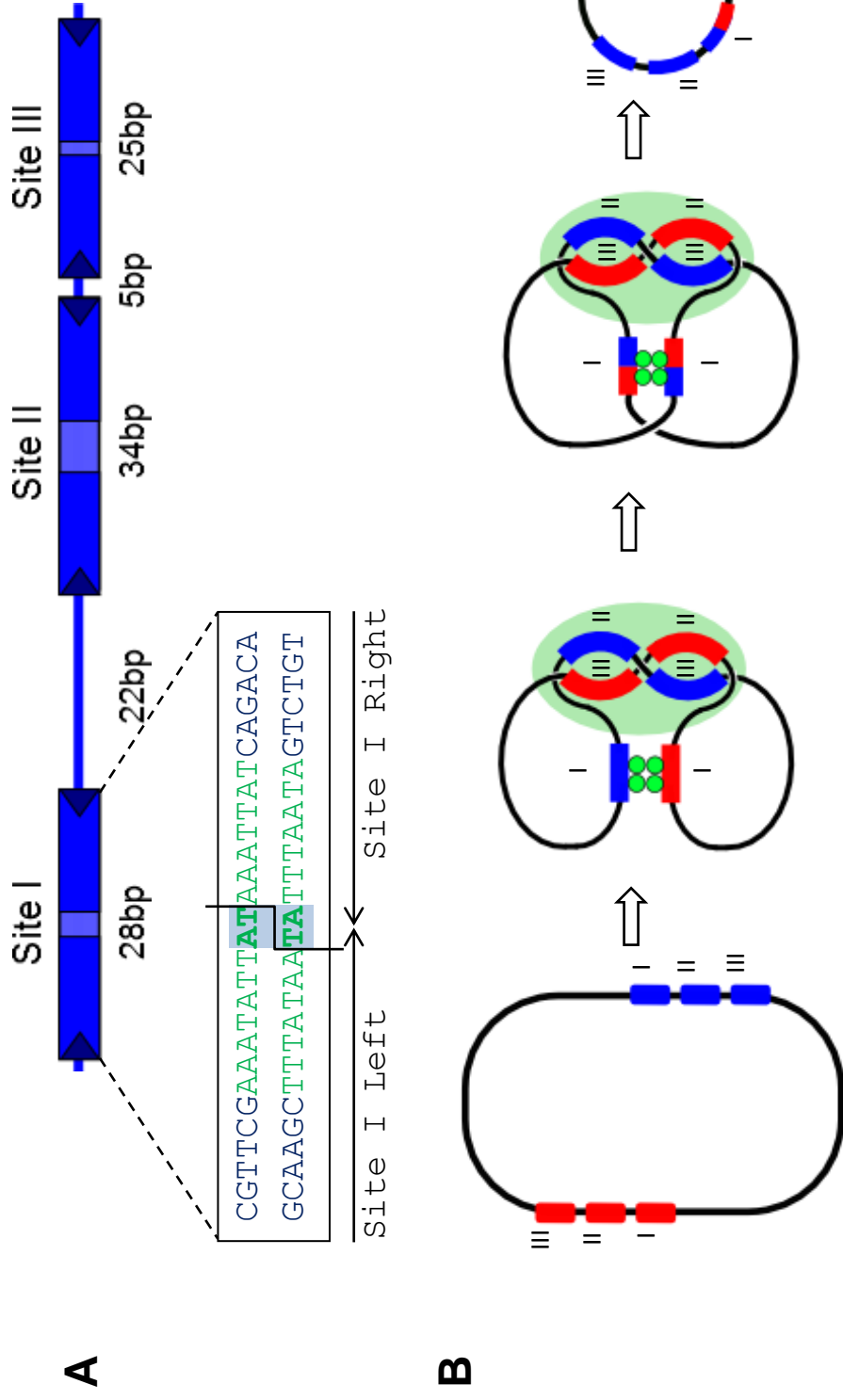
A) The structure of the Tn3 transposon, containing the transposase (red), resolvase (green) and  $\beta$ -lactamase (blue) genes and the *res* site (white box with black triangle). Transposase is expressed and the inverted repeat ends (purple) are recognised by the transposase, leading to replicative transposition and resulting in co-integrate formation. B) The transposition pathway of the replicative Tn3 transposon, transposase catalyses transposition of the Tn3 transposon, causing co-integrate formation. Tn3 resolvase resolves the co-integrate, yielding two copies of the transposon..



**Figure 1.6: Structure of a  $\gamma\delta$  resolvase dimer bound to DNA**

(A) Crystal structure of  $\gamma\delta$  resolvase, showing the N-terminal catalytic domain (green) connected to the E-Helix (blue), linked by an extended linker region (yellow with AT-hook DNA binding motif sitting in the minor groove shown as spheres) to a helix-turn-helix DNA-binding domain (red) bound to DNA (grey). The position of the catalytic nucleophile S10 is shown. Image created using 1GDT crystal structure from PDB (Yang & Steitz 1995).

(B) Cartoon highlighting the functional regions of the structure and showing the 2,3' interface (dark green oval) between the catalytic and regulatory subunits, dimer interface (dashed line) and synaptic interface (dotted line).



**Figure 1.7: Synapsis and strand exchange by Tn3 resolvase**

**A)** Diagram showing the structure of the Tn3 *res* site, illustrating the crossover site (site I) at which double-strand cleavage takes place and the two regulatory sites (sites II & III) which are required for formation of the synapse. Triangles indicate imperfect inverted repeat motifs. The sequence of site I, with the crossover site (black lines, shaded), central 16bp (green) and the outer 6bp recognised by the helix-turn-helix DNA binding domain (dark green) shown. **B)** Cartoon illustrating assembly of the synapse and the products of recombination. Resolvase, illustrated as green circles (illustrated at site I only here), binds as a dimer at each of the *res* sub-sites (I, II & III). The formation of regulatory complexes at the accessory sites (II & III) facilitates the formation of the catalytic tetramer at site I. The product of recombination is a 2-noded catenane.

wild-type (WT) system promote the pairing of resolvase dimers bound at site I. Interactions between the sites promote stability within the structure and regulate recombination. Interactions between site I and site III create a dimer-dimer interaction which promotes catalysis at site I. Interactions between the two *res* sites trap three negative supercoils as can be seen in figure 1.7, thus storing energy for use in the recombination reaction (Stark and Boocock 1995; Burke et al 2004; Grindley et al. 2006).

### **1.3.ii The structure of Tn3/ $\gamma\delta$ resolvase**

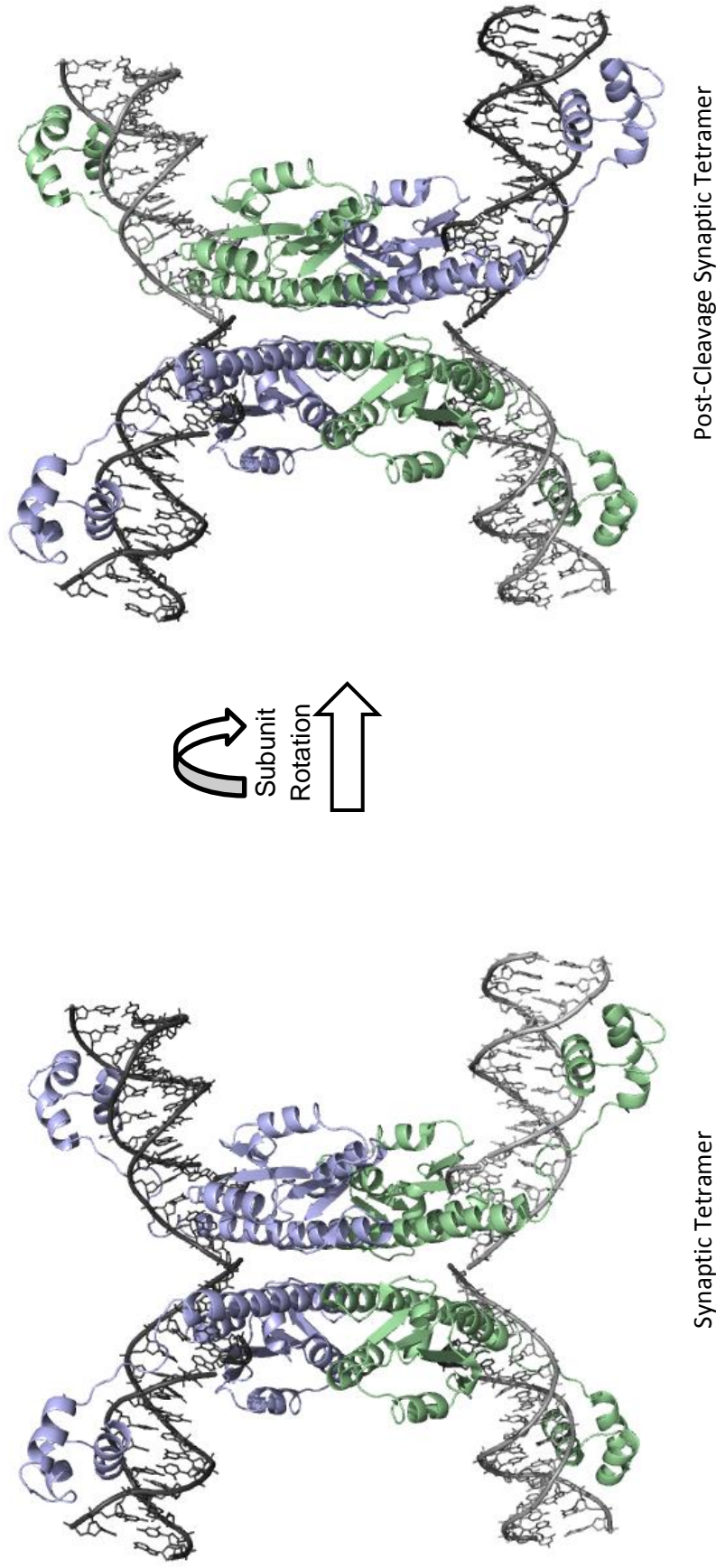
Much of what is known about the domain structure and function of the serine recombinases has come from the archetypal  $\gamma\delta$  resolvase, which has very similar sequence to Tn3 (78.5% sequence identity). The structure of  $\gamma\delta$  resolvase and the related Tn3 and Sin recombinases have been revealed by a number of crystal structures (Yang and Steitz 1995; Li et al. 2005; Mouw et al. 2008; Keenholz et al. 2011). These serine recombinases have a structure consisting of an N-terminal catalytic domain ( $\gamma\delta$  residues 1-100), an extended E-helix ( $\gamma\delta$  residues 101- 120) and E-helix arm ( $\gamma\delta$  residues 121-137) connected by a short linker ( $\gamma\delta$  residues 138-148) to the C-terminal DNA binding domain as shown in figure 1.5. The catalytic domain is made up of four  $\alpha$ -helices ( $\alpha$ A,  $\alpha$ B,  $\alpha$ C and  $\alpha$ D) with  $\beta$ -sheet between  $\alpha$ D and the other  $\alpha$ -helices. Residues R2 and E56 within the catalytic domain are involved in the 2-3' interface between the catalytic site I subunits and the regulatory subunits at site III for regulation of the WT synapse (Sarkis et al. 2001). Residues Y6, R8, S10, D36, R68 and R71 within the catalytic domain are highly conserved and are important for catalysis, as mutation of these residues is detrimental to catalysis by Tn3 at site I (Olorunniji and Stark 2009).

The E-helix is involved in formation of dimer interactions, with hydrophobic residues forming interactions between the E-helices of each subunit, stabilising the dimer structure and positioning the catalytic domain relative to

the DNA. The E-helix remains structured even in the absence of DNA (Rice and Steitz 1994). The 'arm' region of the E-helix is disordered in the absence of DNA. This region forms interactions within the minor groove of the DNA and assists in positioning of the CTD (figure 1.5). The extended linker region ( $\gamma\delta$  residues 138-148) contains the conserved 'GR' A-T hook motif (amino acids 141 and 142) and forms interactions within the minor groove at a conserved AT base pair. The DNA-binding domain has three  $\alpha$ -helices ( $\alpha$ F,  $\alpha$ G and  $\alpha$ H) which surround a conserved hydrophobic core. The  $\alpha$ G and  $\alpha$ H helices form a helix-turn-helix motif which confers DNA recognition for the motifs at each end of *res* site I and the regulatory sites II and III via hydrogen bonds and hydrophobic interactions with specific nucleotides. Hydrogen bonds and salt bridges are formed between phosphate residues in the DNA backbone and residues R148, S162, A171, T174, Y176 and K177. This acts to anchor the CTD within the major groove. Further base-specific interactions are formed between the DNA and residues R172, S173, Y176 and K177.

Two models were proposed for the synaptic interactions of serine resolvases and DNA at the crossover sites; namely, the 'DNA in' and 'DNA out' models. The 'DNA in' model proposed that the DNA is held at the inside of the complex with the resolvase subunits on the outside, the 'DNA-out' model proposed that the resolvase subunits are in the centre, with the DNA on the outside. The DNA-out arrangement was determined by Nöllmann and colleagues (Nöllmann et al. 2004) crystal structures obtained for  $\gamma\delta$  resolvase and Sin resolvase also show that this interaction fits with the 'DNA out' model (Burke et al 2004; Li et al. 2005; Rowland et al. 2009). The currently favoured model for the mechanism of the recombination reaction at the crossover site is the 'subunit rotation' model. This model, suggested by biochemical and topological data and supported by crystal structure data, hypothesises that during recombination the DNA strands are cleaved and then one half of the intermediate protein-DNA complex rotates 180° relative to the other at the flat interface of hydrophobic residues (Grindley et al. 2006) which can be seen at the E-helix interface in crystal structures of recombinase such as that of  $\gamma\delta$  resolvase, as shown in Figure 1.8.





**Figure 1.8: subunit rotation model for strand exchange by serine recombinases**

In this model, the pre-synaptic dimer undergoes conformational changes, facilitating the formation of a synaptic tetramer. The DNA is cleaved on both strands simultaneously and the subunits rotate  $180^\circ$  at the flat interface formed by interactions of the E-helices, followed by re-ligation of the strands. Picture made using the  $\gamma\delta$  resolvase crystal structure 1ZR4 (Li et al. 2005)

Sin resolvase is a member of a group of serine recombinases which are encoded by large *Staphylococcus* plasmids. It was initially identified as a member of the family by Paulsen and colleagues based on sequence analysis (Paulsen et al 1994); its natural role is to reduce plasmid multimers to monomers. This increases plasmid stability, as monomers are maintained at higher copy number in cells than multimers, thus reducing the risk of the plasmid being lost. The resolution product produced by Sin recombinase, like that of Tn3, is a pair of catenated DNA circles.

Sin resolvase is structurally very similar to the previously described Tn3/ $\gamma\delta$  resolvases.  $\gamma\delta$  resolvase has 31% sequence identity to Sin. The crystal structure of DNA-free Q115R Sin (Keenholtz et al. 2011) shows the structure of the catalytic domain and E-helix (residues 1-124); however the C-terminus is not revealed. No crystal structure currently exists showing DNA-bound Sin at site I, although a crystal structure is available which reveals a DNA-bound tetramer at site II (figure 1.9) (Mouw et al. 2008). The structure of Sin will be further discussed in section 1.4.ii.

#### **1.4.i Recombination by Sin**

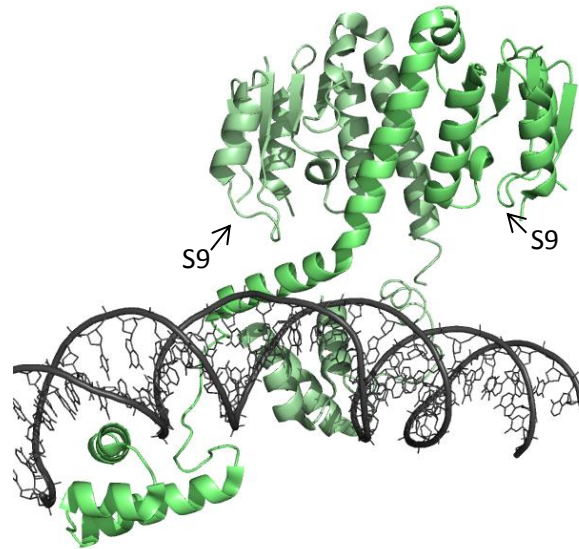
Sin has an 86 bp res site, (*resH*) with a 30 bp catalytic site I, separated by a 31 bp spacer containing a HU binding site from its 25 bp site II (figure 1.10). Site I is made up of a head-to-head repeat sequence, whereas site II is arranged in a head-to-tail repeat. Sin-mediated recombination requires the binding of a DNA bending protein such as HU from *Staphylococcus aureus* or HBSu from *Bacillus subtilis* (Rowland et al 2005, Rowland et al 2006; Grindley et al 2006). The assembly of the Sin synaptosome, like that of Tn3 resolvase, traps three negative supercoils. The strand-exchange reaction is chemically isoenergetic and thus has no necessity for a high energy co-factor. Activated Sin mutants such as I100T and Q115R allow interactions of the catalytic

**A**

CGTATGATTAGGGTGTATATTAATT  
GCATACTAATCCACATATAATTAA

Site II a

Site II b

**B****C**

**Figure 1.9: Structure of a Sin resolvase bound to DNA (*resH* site II)**

**(A)** Sequence of *resH* site II **(B)** A dimer of Sin resolvase bound to the regulatory site II. Site II, unlike site I, has a head to tail orientation of the Sin CTD-binding motifs. Synapsis at site II is mediated by an interface between the DNA-binding domains, as shown in the tetramer structure **(C)**. Position of catalytic serine (S9) is indicated. Images were created using the 2R0q crystal structure (Mouw et al 2008).



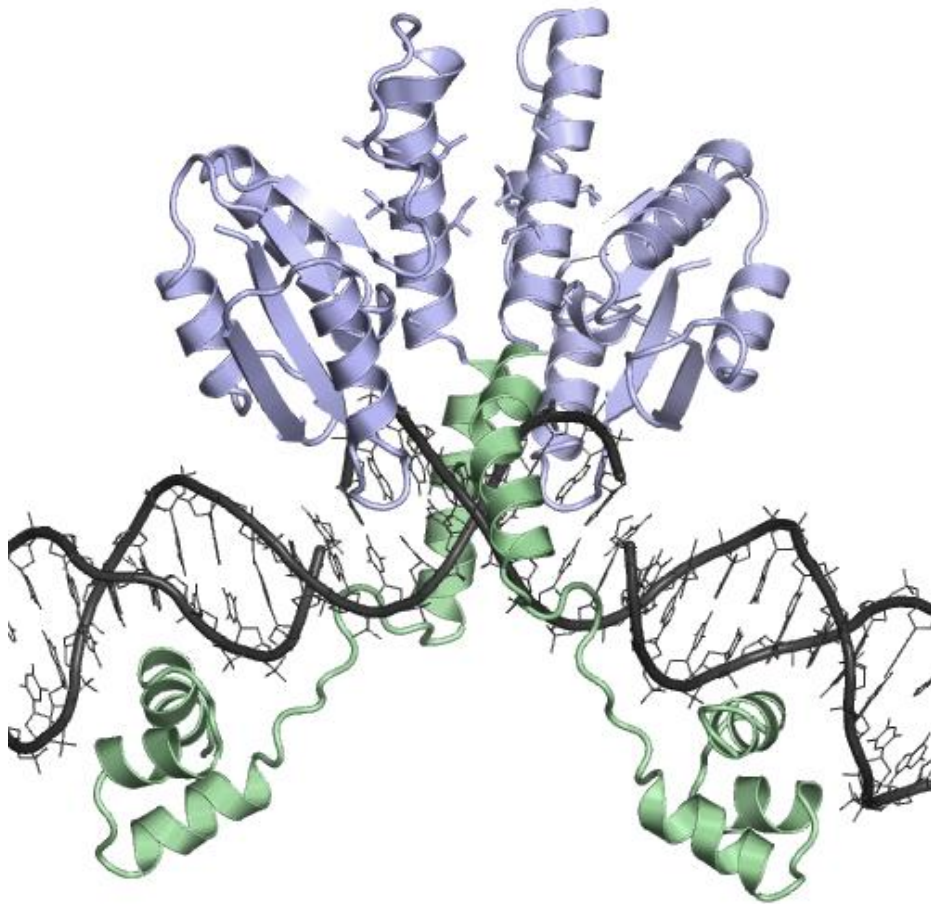
tetramer at Site I to be studied in the absence of the accessory sites and DNA-bending protein (Rowland et al 2005; Mouw et al 2010). The work carried out in this study utilises the Q115R Sin mutant.

#### **1.4.ii The structure of Sin resolvase**

Crystal structure data for the activated mutant of Sin (Q115R) (Keenholtz et al 2011) revealed a tetrameric architecture of catalytic domains similar to that seen in the structure of  $\gamma\delta$  resolvase (Li et al 2005). Sin shares the catalytic domain structure of four  $\alpha$ -helices (A-D) and  $\beta$ -sheet structure connected by a short linker to an extended E-helix as described for  $\gamma\delta$ /Tn3 resolvase in section 1.3.ii. One of the  $\beta$ -sheet structures in  $\gamma\delta$  resolvase is replaced in Sin with a further  $\alpha$ -helix, termed D'. This D'-helix is poorly ordered in the crystals, indicating that it has flexibility and supporting the hypothesis that this acts as a 'hinge', permitting the E-helices to move relative to the catalytic domains. Residues F52 and R54 of the catalytic domain form the Sin equivalent of the  $\gamma\delta$  resolvase 2-3' interface between sites I and III (Mouw et al. 2008) which connects dimers of resolvase and is essential for wild-type resolvase to assemble the synapse of two *res* sites.

This Sin structure does not reveal the DNA-binding domain structure. However, the homology shared by Sin and Tn3/ $\gamma\delta$  resolvases means that the structure formed by Sin at site I can be extrapolated, by docking of the Sin structure (Keenholtz et al. 2011) to the DNA-bound CTDs of the  $\gamma\delta$  structure (Yang and Steitz 1995). Such a model is shown in figure 1.11.

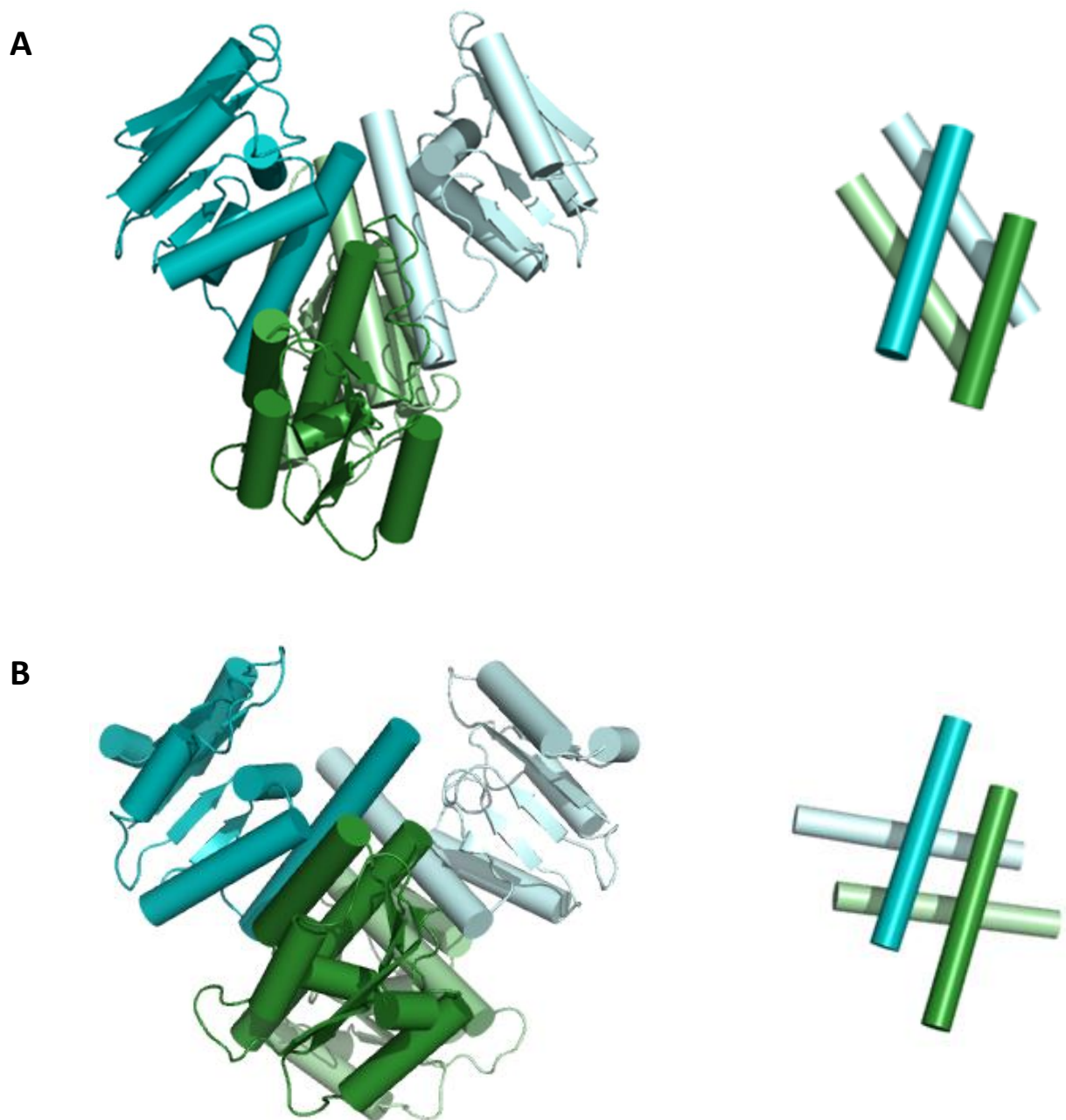
The activated Sin Q115R crystal structure shows the tetramer in a different rotational state than has previously been seen in structures of other resolvases. The Sin E-helix has an insertion at the N-terminal end, making it one turn longer relative to the  $\gamma\delta$  resolvase structure and altering the angle of the E-helices of Sin relative to  $\gamma\delta$  resolvase (see figure 1.12), as a result of which, the cutting dimers of Sin are held further apart than those of  $\gamma\delta$  resolvase.



**Figure 1.11: Model of Sin activated mutant dimer bound to site I**

This model shows Sin (blue) residues 1-124 (crystal structure 3PKZ) docked by rigid body docking to  $\gamma\delta$  resolvase CTD (green) (crystal structure 1GDT) bound at site I. This image shows half of a hypothetical catalytic tetramer, the Sin shown is an activated mutant from a tetramer structure. Image created using 3PKZ and 1GDT crystal structures from PDB (Keenholtz et al 2011, Yang and Steitz 1995).





**Figure 1.12: The E-helices of Sin and  $\gamma\delta$**

**(A)** Catalytic domains and E-helices of DNA-free Sin tetramer (left) and E-helices only (right) (3PKZ). **(B)** Catalytic domains and E-helices of DNA-bound  $\gamma\delta$  tetramer (left) and E-helices only (right) (1ZR4). The altered rotational state of Sin relative to  $\gamma\delta$  leads to an altered angle of E-helices. Figure made using 3PKZ and 1ZR4 crystal structures (Keenholtz et al 2011, Li et al 2005)

A crystal structure of the regulatory Sin tetramer bound to two copies of site II (Mouw et al. 2008) (illustrated in figure 1.9) showed at high resolution the  $\alpha\beta$  catalytic N-terminal domain, helix-turn-helix (H-T-H) DNA-binding C-terminal domain and the intervening extended E-helix. This regulatory structure reveals asymmetry in the Sin dimers. This asymmetry permits the docking of the CTDs to the head-to-tail DNA sub-sites in the configuration seen in the crystal structure. In this structure, the catalytic serines of each monomer are partially blocked from the DNA by the positioning of the E-helix of the other monomer, so that the protein is catalytically inactive at site II.

The E-helix region of resolvase is involved in formation of dimer interactions, with interactions between the E-helices stabilising the dimer structure and positioning the catalytic domain relative to the DNA. The modularity of resolvase offers a way to study the effect of the E-helix on synapsis and DNA binding, as the catalytic domain can be removed, yielding a truncated variant. Such truncated variants termed ‘tadpoles’ shall be discussed in detail in chapter 3.

## **1.5 Genetic applications of site specific recombinases**

Targeted modification of the genome has a variety of applications in research, medicine, and biotechnology. Site-specific recombinases are important tools in these areas. Site-specific recombinases allow the precise integration of open reading frames encoding pharmaceutically relevant proteins into specific gene loci in cell lines and transgenic animals (Kolb 2004).

Since site specific recombination typically occurs between DNA sequences which have evolved very specifically along with a resolvase enzyme to obtain optimum efficiency, they have very sequence specific targeting. In order for recombinases to be useful as tools for genome editing, it is necessary to alter



the protein to change its natural sequence selectivity. This can be achieved by directed evolution to achieve recognition of pseudo-sites or by domain swapping for recognition of different sites. Directed evolution has been used successfully with the large serine recombinase  $\phi$ C31 integrase, altering its specificity to target pseudo *attP* sites in the human genome (Thyagarajan et al. 2008).

Due to the modular nature of resolvase, the DNA-binding and catalytic domains can be separated into distinct domains. This means that the DNA-binding domain can be removed and swapped for that of another protein to give targeting to a different site. Domain swapping experiments, creating chimeric proteins with the catalytic activities of the chosen catalytic domain and the DNA-targeting of the chosen DBD have been carried out with members of the serine recombinase family. Schneider et al. demonstrated that the DNA-binding domain of Gin invertase could be replaced with that of Tn3 resolvase, creating a chimeric protein which interestingly was able to function as a resolvase, but not as an invertase (Schneider et al. 2000). Domain swapping experiments have also been carried out with Tn3 resolvase, in which the DNA binding domain of Tn3 resolvase was replaced with a zinc finger DNA binding domain (Akopian et al. 2003, Prorocic et al. 2011) additionally, the Gin invertase DBD was replaced with that of a transcription activator like effector (TALE) binding domain (Mercer et al. 2013). Zinc finger recombinases (ZFRs) and transcription activator like effector recombinases (TALERS) will be discussed in more detail in sections 1.7.iv and 1.7.v respectively.

This recombinase modularity also means that the DNA-binding domain of resolvase is able to bind in the absence of the catalytic domains, a property which allows closer examination of the E-helix domain, as will be discussed in chapter 3.

## 1.6 Other methods of genome editing

Methods of genome editing such as alteration of gene expression by chromosomal integration of DNA by homologous recombination are limited and can have off-target effects. Integration at random loci in the absence of targeting is inefficient and the gene may be subjected to locus-specific epigenetic effects which may down-regulate or silence it (Pannell & Ellis 2001). Insertion at certain positions has the potential to disrupt local genes and cause detrimental effects. Gene knockout via HR can be prone to off-target effects, and can be inefficient dependent on cell-type (Smith 2001). Additionally, as the HR system is essential for cellular functions, its use for engineering purposes may negatively impact on local genes and have disruptive effects on local gene expression (Akopian and Stark 2005).

Addition of genes into mammalian cell lines is of interest for medicine, synthetic biology, engineering of cell systems and gene function studies. Some methods of chromosomal gene insertion involve random integration within the genome. These random insertions can have potentially damaging effects, as seen in clinical gene therapy trials, in which the insertion of a viral vector containing a transgene with a strong viral promoter into the genome proximal to proto-oncogenes has led to de-regulation and cellular clonal expansion (Ott et al. 2006, Fischer et al. 2010, Cavazzana-Calvo et al. 2010).

Transposon-derived insertion vectors like Piggybac and Sleeping Beauty are however useful for other applications such as identification of potential proto-oncogenes in murine models, although their uses for gene therapy are limited by their non-specific insertions (Collier et al. 2005, Ivics et al. 2007, Rad et al. 2010).

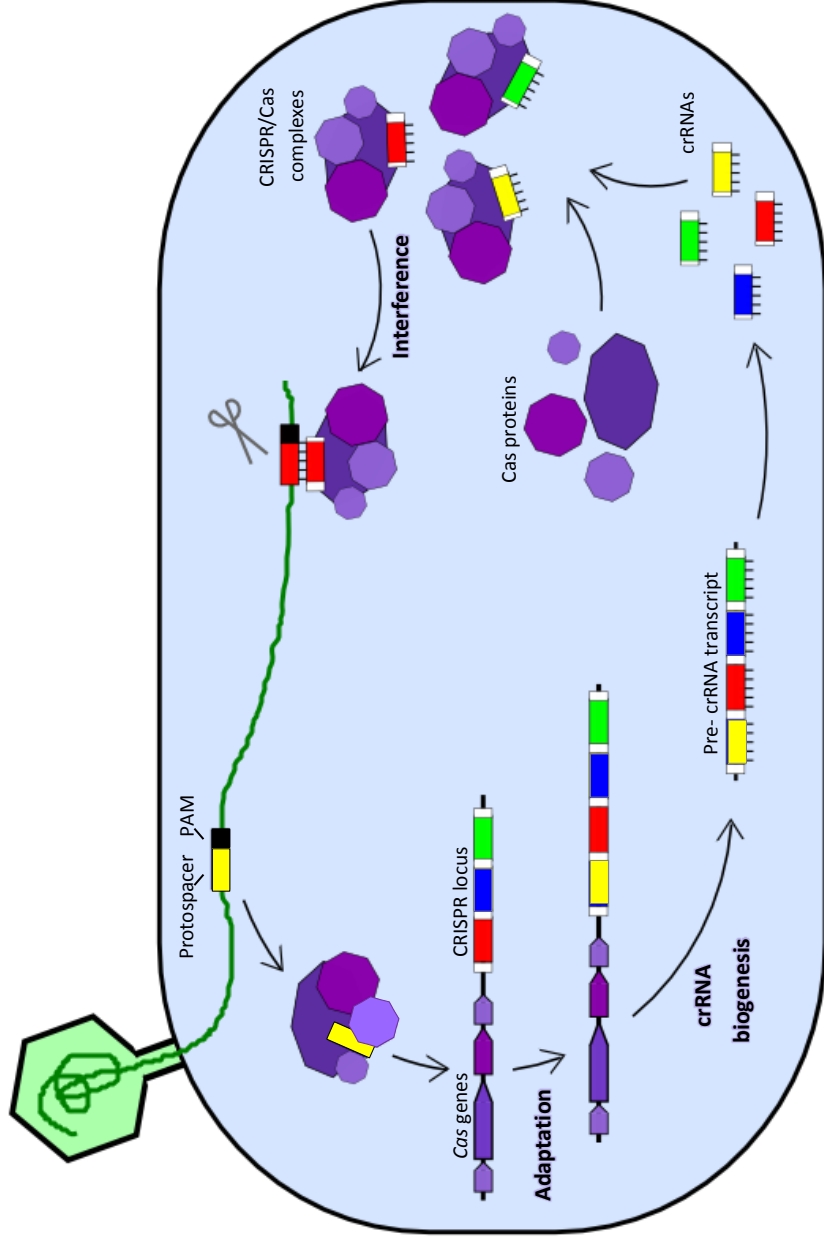
The focus of this work is the study of DNA targeting by members of the site-specific serine recombinase family. Previous work carried out by the Stark group led to the development of zinc finger recombinases (ZFRs). This work will discuss the creation of a similar chimeric protein, a TALE recombinase

(TALER). However, it is important to note that there are several other methods of genome editing available, including the zinc finger nucleases (ZFNs), the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas 9 system and Transcription Activator-Like Effector Nucleases (TALENs), which shall be discussed briefly.

### **1.6.i The CRISPR/Cas System**

An emerging alternative to the zinc finger and TALE DNA targeting mechanisms is the CRISPR/Cas system. The CRISPR system is a bacterial acquired immunity system protecting against invading foreign DNA by RNA-guided DNA cleavage (Barrangou et al. 2007).

There are three main types of CRISPR/Cas system, with many more sub-types. The most commonly utilised of these is the type II system. There are three main steps in the type II system; adaptation, biogenesis and interference. During adaptation, new spacers are incorporated into the CRISPR locus; these spacers are acquired from protospacer sequence in invading DNA. This CRISPR locus is made up of direct repeats, separated by the spacer sequences. Biogenesis is the process of transcription of this CRISPR locus, forming pre-crRNA primary transcript, which is then processed to form short crRNAs. These crRNAs associate with CRISPR associated (Cas) proteins and these CRISPR/Cas complexes provide immunity against infection by plasmids and bacteriophages. Upon recognition of a complementary protospacer and PAM DNA sequence in the invading DNA, the CRISPR/Cas complex blocks replication of foreign DNA and initiates DNA cleavage (Figure 1.13) (Samson et al. 2013).



**Figure 1.13: The type II CRISPR/Cas system**

Foreign DNA is introduced into the cell, and spacer acquisition occurs. The protospacer and adjacent protospacer adjacent motif (PAM) are recognised and a new spacer is incorporated into the CRISPR locus in a process known as adaptation. Biogenesis of crRNAs then takes place and a pre-crRNA transcript is produced, which is processed to form crRNAs. Complexes form containing CRISPR associated (Cas) proteins and a crRNA. These complexes then recognise protospacer and PAM sequences on invading DNA and provide immunity by blocking replication of and initiating cleavage of identical sequences in the foreign DNA in a process known as interference (figure adapted from Samson et al 2013).

The CRISPR/Cas system has great potential for targeting, as re-design of the crRNA can allow targeting of Cas cleavage to a desired site. Similar to zinc fingers and TALEs, Cas9, the most commonly used type II Cas protein has two nuclease domains (RuvC and HNH) and as been engineered to act as an endonuclease and as a nickase for RNA-guided site-specific DNA cleavage (Cong et al. 2013).

### **1.6.ii Zinc Finger Nucleases**

Engineered nucleases, using the cleavage domain of a nuclease and a sequence-specific DNA-binding domain, have been developed for 'genome editing', creating targeted double strand breaks which stimulate cellular DNA repair mechanisms such as Homology Directed Repair (HDR) and Non-Homologous End Joining (NHEJ), with the targeting specificity conferred by zinc finger or TALE DNA binding domains.

The zinc finger Cys<sub>2</sub>-His<sub>2</sub> domain is one of the most common motifs in eukaryotes. Each finger is made up of approximately 30 amino acids in a conserved  $\beta\beta\alpha$  configuration. 3 bp of DNA in the major groove of the DNA is recognised by residues on the surface of the  $\alpha$ -helix, with each finger therefore giving specificity for a specific triplet of bases (Beerli and Barbas 2002).

Due to the modular nature of zinc finger proteins, different sequence selectivity can be designed by swapping the fingers. Artificial zinc finger domains with the ability to recognise 9-18 bp of DNA have been created (Dreier et al. 2001, Dreier et al. 2005). The development of such artificial zinc finger domains greatly increases their specificity, as 18 bp of DNA recognition theoretically confers specificity for 68 billion base pairs of DNA ( $4^{18}$ ), meaning that genomic targeting became possible.

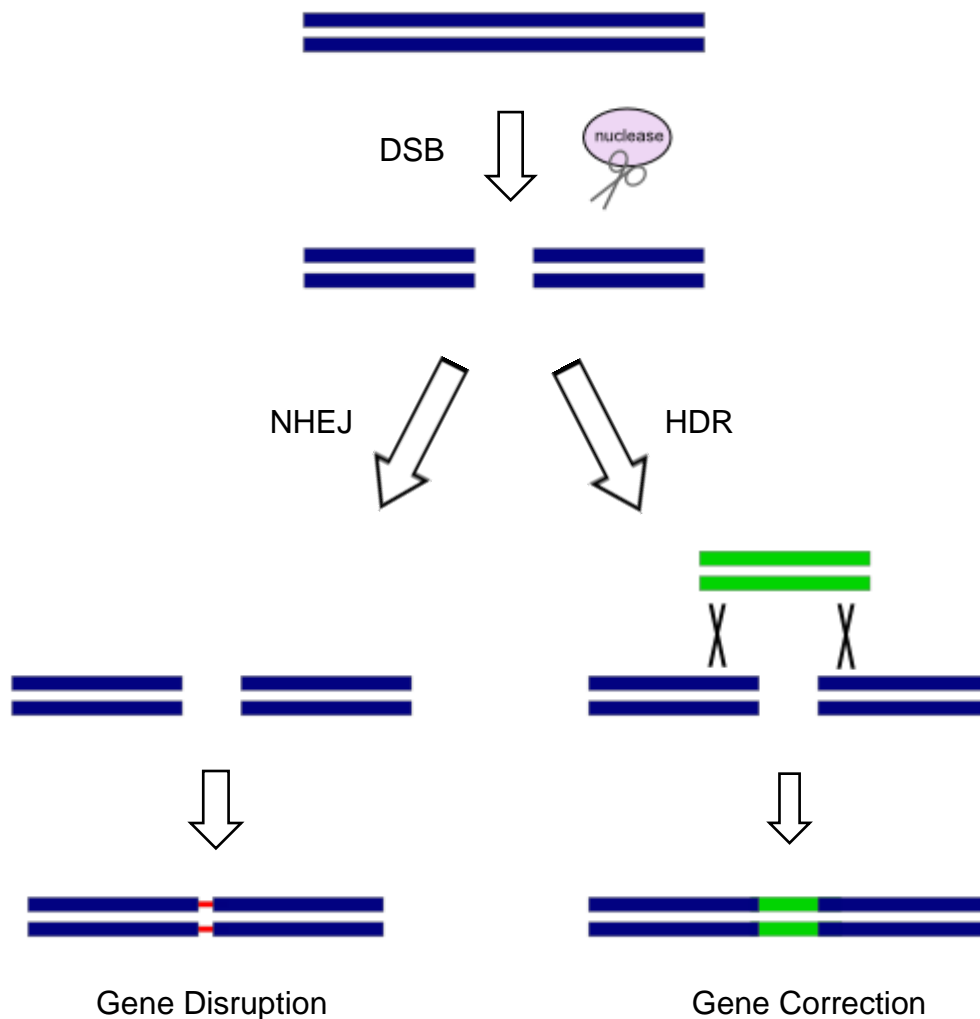
Zinc finger assembly can now be carried out using a zinc finger library (Beerli and Barbas 2002) or by other methods, such as the selection-based randomised approach, known as Oligomerized Pool Engineering (OPEN) (Maeder et al. 2008).

Traditionally, the process of gene replacement, inactivation or insertion has utilised homologous recombination techniques. However, homologous recombination is carried out inefficiently in mammalian cells. Chimeric nucleases, such as the ZFNs have been used to increase this efficiency by targeting double-strand breaks to a specific locus (Moehle et al. 2007).

The double-stranded cleavage of DNA by site specific nucleases can lead to HDR, which can be used for gene correction or NHEJ (Figure 1.14). NHEJ is error-prone by nature and as such introduces small insertions or deletions. This can cause frame-shift mutations. In order to negate the effects of this, nickases, nucleases engineered to nick the DNA instead of causing double-strand breaks have been developed, which stimulate HDR without promoting NHEJ (Kim et al. 2012 and Gaj et al. 2012).

### **1.6.iii TALE Nucleases (TALENs)**

An addition to the available DNA-binding domains for creation of chimeric nucleases and recombinases is the Transcription Activator-Like Effector (TALE) family of proteins. TALEs are found in nature in plant pathogens (*Xanthomonas* sp.) where they act as transcription factors to directly modulate host gene expression. They have DNA-binding domains which consist of 33-35 amino acid repeats (typically 34 amino acids), each of which confers recognition to a single base pair of DNA (figure 1.15) (Mak et al. 2012; Deng et al. 2012).



**Figure 1.14: Possible outcomes of nuclease DNA targeting**

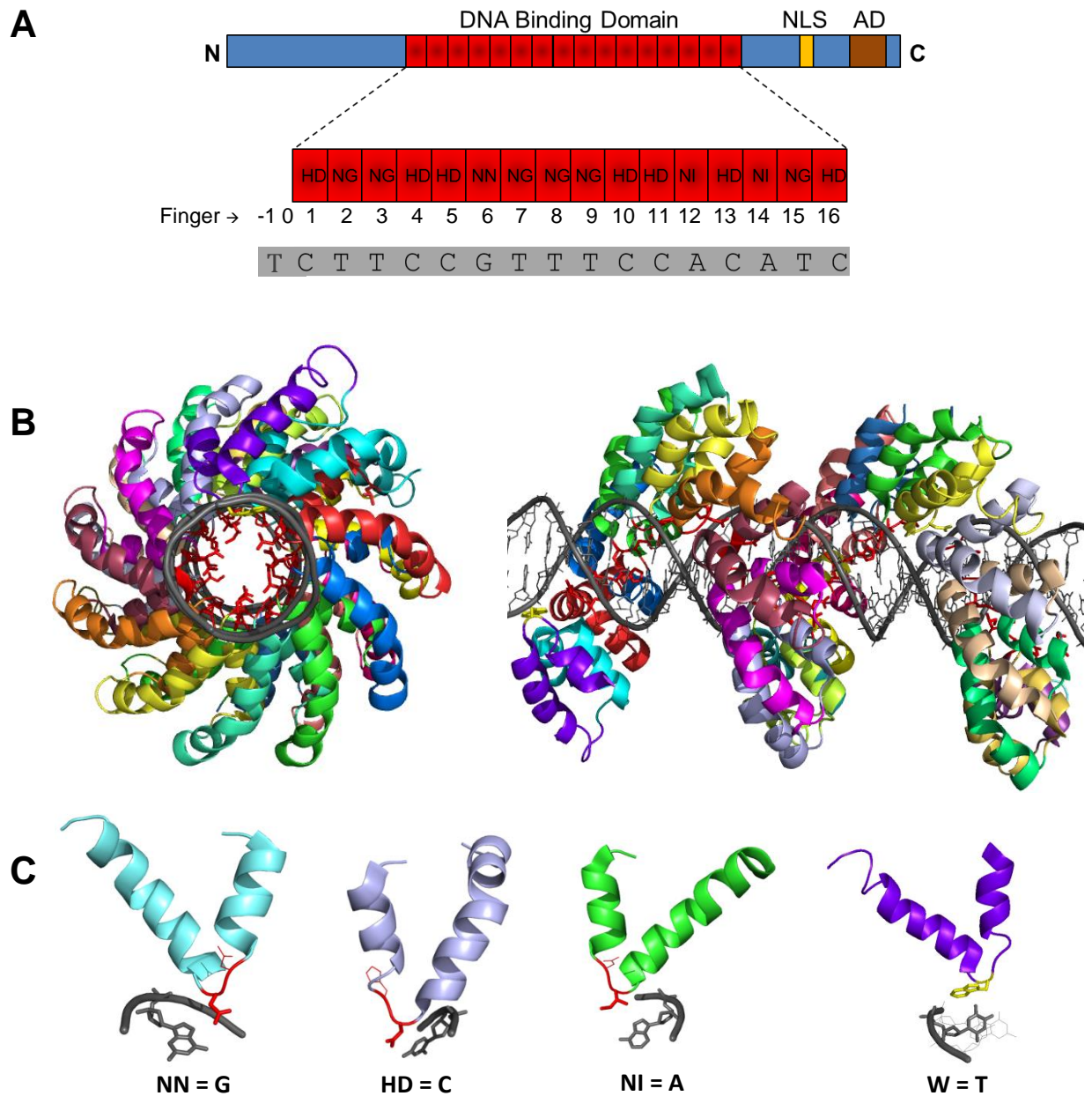
Nuclease-based DNA targeting can either lead to Non-Homologous End Joining (NHEJ) or Homology Directed Repair (HDR). NHEJ naturally leads to small insertions or deletions often causing gene disruption. HDR in nature uses a DNA molecule with sequence homology to repair double strand breaks. Introduction of a plasmid with flanking homologous regions can be used for gene correction.

The amino acids at positions 12 and 13 of the repeats form the repeat variable di-residue (RVD), a hyper-variable region which is responsible for sequence recognition. The amino acid at position 13 of each repeat is involved in recognition of the DNA, while the amino acid at position 12 is involved in stabilising the protein structure by interactions with the neighbouring repeat (Streubel et al. 2012). The sequence which is recognised by these RVDs can be altered by altering the order of these RVD containing repeats.

The RVD nucleotide recognition code has been studied, and as such, the nature of DNA binding by RVDs is now better understood. The four most common RVDs found in nature are: HD, NG, NI and NN (which recognize C, T, A and G/A respectively). In 2009, studies of the TAL recognition code identified the sequence selectivity by common and rare RVDs highlighting the sequence recognition selectivity by the repeats (Boch et al. 2009, Moscou & Bogdanove 2009).

The complete modularity of these TAL repeat domains is unique amongst the commonly studied DNA-binding domains, which can have positional constraints for modular assembly, as in other cases such as zinc finger domains, neighbouring fingers can have an effect on sequence recognition of the adjacent finger (Miller et al. 2008), meaning that certain sequences are more efficiently bound than others, and some sequences cannot be targeted. Zinc finger domains are costly and time consuming to develop and have sequence limitations which the TALs do not have, with certain sequences for which no compatible zinc finger binding domains are available. The TALEs by comparison are relatively simple to construct and to customise, making them potentially very useful as tools for genome engineering and for synthetic biology purposes (Zhang et al. 2011; Mussolino et al. 2012).





**Figure 1.15: TALE Structure**

**(A)** TAL Effector domain organisation, showing the DNA-binding domain repeats (red boxes) with RVDs from the TALE DBD used in this study and the DNA sequence recognised by it (grey), the nuclear localisation domain (yellow) and activation domain (brown) which are required for natural function. **(B)** Crystal structure of a TALE PthXo1; (PDB ID: 3UGM) bound to DNA, with repeat finger highlighted in different colours. The fingers can be seen to wrap in a spiral around the DNA with the RVD positions 12 and 13 at the centre of the structure interacting specifically with the DNA **(C)** Individual TAL fingers bound specifically to a specific nucleotide, according to the repeat variable di-residue (RVD) present at positions 12 and 13 of the repeat., and recognition of preceding 'T' by the tryptophan (w) residue in finger -1 (yellow).

Although the TAL effectors are largely free from many of the constraints posed by other DNA-binding domains such as zinc fingers, they do have one limitation, namely, the necessity of the recognition sequence to have a 'T' base 5' of the recognition sequence. This highly conserved nucleotide is believed to interact with the TAL '0' and '-1' repeats, improper repeat fingers N-terminal of the RVD-containing canonical repeats, as demonstrated by Mak and colleagues, whose 2012 crystal structure of the PthXo1 TALE shows that a tryptophan in the '-1' finger N-terminal of the near-identical RVD containing repeats is located near enough to this 5' T base to interact with it (Mak et al. 2012).

The 5' T has been shown to be necessary for efficient TALE binding and adds an undesirable restriction to the use of TALE domains for site-specific DNA targeting. Tsuji and colleagues recently examined the possibility of removing this requirement for the 5' T by targeted mutagenesis of the '-1' finger of the TALE (Tsuji et al. 2013). They made single point mutations, changing the tryptophan in the '-1' finger to other amino acids, and carried out targeted mutagenesis of four amino acids in the loop region of this finger. They showed that through directed evolution of this region, they were able to remove the T constraint from the 5' end. Their work also indicated that the tryptophan is important to the binding of the T, as single point mutations of this residue did not allow the binding of sequences with no 5' T, but did prevent efficient binding to the 5' T substrate.

TALEs, like zinc fingers have been used to construct hybrid nucleases (Christian et al. 2010), transcriptional activators (Mussolino et al. 2011) and recombinases (Mercer et al. 2012). As TALE modules have single base pair recognition, they have greater flexibility for construction; however, initially their highly repetitive nature was problematic for synthesis and assembly. Methods have now been developed which allow easy assembly of custom TALE arrays (Cermak et al. 2011, Reyon et al. 2012, Briggs et al. 2012, Schmid-Burgk et al. 2013). The development of new, efficient TALE

assembly methods has facilitated the assembly of libraries of TALE arrays, such as the recently created library of TALENs which is able to target 18,740 human protein coding genes (Kim et al. 2013).

ZFNs and TALENs have been successfully used for therapeutic applications, such as the use of ZFNs to correct mutations associated with Severe Combined Immunodeficiency (SCID), sickle cell disease and haemophilia-B (Urnov et al. 2005, Sebastiano et al. 2011 and Li et al. 2011) and the use of TALENs in targeted correction of the sickle cell disease-causing mutation in the  $\beta$ -globin gene in human cells (Sun et al. 2012).

#### **1.6.iv Zinc finger recombinases**

Zinc finger recombinases (ZFRs) are site-specific recombinases with the catalytic domain from resolvase proteins such as the Tn3 and Sin resolvases, fused to the DNA binding domain of a zinc finger protein such as the mouse transcription factor Zif268. Designer recombinases such as ZFRs, like the ZFNs previously discussed, have great potential for use in biotechnology, gene therapy and experimental genetics (Urnov et al. 2005, Sebastiano et al. 2011 and Li et al. 2011).

The development of 'activated' serine recombinases discussed in sections 1.2 and 1.4, which act on their catalytic site without the requirement for the accessory site proteins have been essential for the creation and success of chimeric proteins such as the ZFRs. ZFRs with activated Tn3 resolvase and Gin invertase domains have been shown to be functionally active with in human cell lines (Gordley et al. 2007 & 2009, Gersbach et al. 2011, Gaj et al. 2013). Such experiments indicate that despite the limitations which have been discussed, there is a great deal of potential for these types of chimeric recombinases for applications such as genomic editing.

Site-specific recombinases have been successfully used as molecular biology tools to specifically carry out integration and excision of DNA sequences. The usefulness of natural recombinases however is limited by the necessity to either rely on the presence of pseudo-sites in genomic DNA or to first introduce cognate recognition sites into the genome prior to expression of the recombinase, for example by HR.

The zinc finger DNA-binding domain (DBD) is a small DNA-binding domain, consisting of three fingers, each of which typically confers recognition to 3 bp of DNA, giving recognition of a 9 bp DNA sequence. ZFRs are chimeric proteins with the DNA binding domain from a zinc finger protein fused to the activated mutant catalytic domain of a site-specific recombinase such as Tn3 resolvase. Such ZFRs have been shown to recombine specially designed 'Z-sites' with the zinc finger binding site flanking the natural crossover site of resolvase *in vivo* and *in vitro* (Akopian et al. 2003).

Chimeric proteins such as ZFRs, which can have their DNA-binding domains altered to bind to many different sequences, reduce this limitation, allowing targeting to be carried out to a wider variety of sequences. Previous work carried out in our lab showed that a ZFR with an activated mutant version of the Tn3 resolvase catalytic domain could be efficiently targeted to specifically designed 'Z-sites' (Akopian et al. 2003, Prorocic et al. 2011).

A limitation of the ZFRs is the specificity for the centre of the target DNA sequence, which is conferred by the resolvase domain. However, directed evolution of recombinases has been carried out to give these proteins activity at sites which resemble, but are not identical to their original cognate site (Sirk et al. 2014). Previously in the Stark lab Tn3-ZFRs were successfully targeted to Z-sites containing the sequences of the bovine  $\beta$ -casein gene, a potentially useful target for gene integration (Proudfoot et al. 2011).

### **1.6.v TALE Recombinases (TALERS)**

ZFRs have been shown to be useful and effective for targeting to certain sites. However they suffer from issues such as a lack of high-affinity recognition of all DNA triplets, imperfect modularity, and difficulty of construction, limiting their usefulness for genome engineering. The emergence of TALEs and the demonstration of their effectiveness in targeting TALENs led to the design and creation of TALE recombinases (TALERS).

TALERS, utilising the specific DNA-binding capabilities of TALEs and the catalytic capabilities of activated recombinase mutants, have the potential to surpass the ZFRs, as the TALE domains do not share the above limitations for sequence selectivity and thus can potentially be targeted directly to more genomic sequences. The creation of a TALER with the catalytic domain of an activated Tn3 resolvase will be discussed in chapter 4.

### **1.7 Project aims**

Site-specific recombination, using a zinc finger or TALE DNA-binding domain fused to the C-terminus of the catalytic domain of a serine recombinase in place of its natural helix-turn-helix domain has the potential to specifically carry out recombination at a desired locus. As the site-specific recombinases carry out all of the steps required for DNA cleavage, rearrangement and re-ligation, they have an advantage over the nucleases, which rely on the cellular repair machinery to repair the double-strand breaks created by the nuclease domain. This gives the recombinases potentially higher activity with fewer off-target effects than their nuclease counterparts. However, in order for recombinase proteins to have optimal functionality, it is important to have a very detailed knowledge of the protein, its structure and functions.

This project examined the structure and functions of members of the serine recombinase family, including the Tn3 and Sin family recombinases, in order to determine which regions of these proteins are involved in processes such as DNA binding, dimer interactions and synapsis. A further goal of this project was the creation of a chimeric TALER with specific activity targeted to a chosen DNA site.

## Chapter 2: Materials and Methods

### 2.1 Bacterial Strains

Both of the *Escherichia coli* strains used in this study are derivatives of the K-12 strain. Their names, genotypes and original sources are given in Table 2.1.

Bacterial Strain	Genotype	Source
DS941	AB1157, <i>recF</i> , <i>lacZ</i> ΔM15, <i>lacI</i> <sup>q</sup>	Summers and Sherratt, 1988
BL21(DE3)[pLysS]	<i>Hsd</i> , <i>gal</i> , ( <i>λcl</i> ts 857, <i>ind1</i> , <i>sam7</i> , <i>ini5</i> , <i>lac</i> <sub>UV5</sub> -T7 gene-1), T7 lysozyme expressing plasmid pLysS	Studier et al 1990

**Table 2.1:** Bacterial strains used in this study

### 2.2 Bacterial Growth Media

Liquid cultures of *E.coli* were grown at 37°C in L-Broth (10 g tryptone, 5 g yeast extract and 5 g NaCl, made up to 1 L with de-ionised H<sub>2</sub>O and adjusted to pH 7.5 with NaOH). Bacteria were also grown on solid media (L-broth with 15 g/L agar).

## 2.3 Antibiotics

Antibiotics were typically made up as 500x or 1000x stocks for storage purposes, shown in Table 2.2.

Antibiotic	Stock Concentration	Final Concentration
Ampicillin (500x)	50 mg/ml in H <sub>2</sub> O	100 µg/ml
Chloramphenicol (1000x)	25 mg/ml in EtOH	25 µg/ml
Kanamycin (1000x)	50 mg/ml in H <sub>2</sub> O	50 µg/ml
Streptomycin (500x)	50 mg/ml in H <sub>2</sub> O	50 µg/ml

**Table 2.2:** Antibiotic stock and working concentrations.

## 2.4 Chemicals

Chemicals used in the assays were purchased from suppliers, as shown in Table 2.3 below.

Chemical	Source
General chemicals, organic solvents and biochemicals	Aldrich/Sigma
Media	Difco
Agarose and acrylamide	Biorad
Radiochemicals	Amersham Biosciences
Restriction enzyme buffers	NEB, Roche, Promega

**Table 2.3:** Chemical reagents utilised in this study and their suppliers.



## **2.5 Gene synthesis**

Synthetic constructs for creation of TALER expression plasmids were designed and ordered from Life Technologies, GeneArt Gene Synthesis.

## **2.6. Oligonucleotides**

All oligonucleotides used in this study for cloning and plasmid construction were ordered from Eurofins Genomics (Table 2.4).

## **2.7. Plasmids**

Plasmids were constructed by restriction digests and fragment swaps from existing plasmids. A list of plasmids created and used is provided in Table 2.5. Construction methods for groups of expression plasmids are outlined in figure 2.1.

Expression plasmids with both high and low levels of expression were used in this study, with high-level expression vectors being used for overexpression of resolvase proteins for purification and use in *in vitro* assays and low-level expression vectors used for *in vivo* experiments.

### **2.7.i High-level resolvase expression plasmids**

All overexpression plasmids constructed are based on pSA1101 (Arnold et al 1999). pSA1101 has a ColEI origin of replication, encodes a kanamycin resistance marker and contains an inducible T7 promoter, which induces overexpression of resolvase cloned in between NdeI and Asp718 restriction sites in the pSA1101 backbone (figure 2.1).

Oligo ID:	Size (bp)	Sequence (5'-3')	Purpose
M13 uni (-43)	23 bp	AGGGTTTTCCCAGTCACGACGTT	Sequencing primer
M13 rev (-49)	24 bp	GAGCGGATAACAATTTACACAGG	Sequencing primer
T7	20 bp	TAATACGACTCACTATAGGG	Sequencing primer
MB67	58 bp	CATGCCGGGAGCTCACTAGTATGTATGATTAGGGTATACCCTAATCATA CAAAGAATT	To introduce symmetrised (To RHS) Sin Site I
MB71	58 bp	CTAGAATTCTTTGTATGATTAGGGTATACCCTAATCATACTAGTGA GCTCCCGG	To introduce symmetrised (To RHS) Sin Site I
MB68	58 bp	CATGCCGGGAGCTCACTAGTATGTATGATTATATTATAATATAATCATAC AAAGAATT	To introduce symmetrised (To RHS) Tn3 Site I
MB72	58 bp	CTAGAATTCTTTGTATGATTATATTATAATATAATCATACATACTAGT GAGCTCCCGG	To introduce symmetrised (To RHS) Tn3 Site I
MB70	58 bp	CATGCCGGGAGCTCACTAGTAAACGGCGGCTTTACGTACCCTAAT CATACAAAGAATT	To introduce symmetrised (To RHS) $\Delta$ Sin Site I
MB74	58 bp	CTAGAATTCTTTGTATGATTAGGGTACGTAAAGCCGCGTTTACTAGTG AGCTCCCGG	To introduce symmetrised (To RHS) $\Delta$ Sin Site I
SEH1-T	54 bp	TATGAACCCATTACTAGATAAATTCATGAAAGACCTAATCATTCAAATATTAGC	To introduce truncation of Sin NTD before residue 101
SEH1-B	56 bp	CATGGCTAATATTTGAATGATTAGGTCTTTCATGAATTTATCTAGTAATGGGTTC A	To introduce truncation of Sin NTD before residue 101

**Table 2.4:** Oligonucleotides used in this research.

Oligonucleotides were ordered from MWG as two sequences; top (T) and bottom (B) which were then annealed prior to cloning.

Oligo ID:	Size (bp)	Sequence (5'-3')	Purpose
SEH4-T	82 bp	TCGAACGAATGAGGGCCGACAGGAAGCAAAGCTGAAAGGAATCAAA TTCGGCCGCCCTTTGCTTTATTCACCGAACGCGAAG	For insertion of Tn3 E-helix 'arm' into Sin
SEH4-B	82 bp	GATCCTTCGCGTTCGGTGAATAAAGCAAAGGGCGGCCGAATTTGATT CCTTTCAGCTTTGCTTCCTGTTCGGCCCTCATTCGT	For insertion of Tn3 E-helix 'arm' into Sin
SEH9-T	20 bp	GATCCTGATCAGCGGTACCG	linker sequence to attach TALE DNA Binding domain to resolvase
SEH9-B	20 bp	CCGGCGGTACCGCTGATCAG	linker sequence to attach TALE DNA Binding domain to resolvase
SEH10a-T	61 bp	CTAGTAAACAGTGGTCGGGAGCCCGGGCACTTGAGGCGCTGCTGA CTGTGGCGGGTGAGCT	linker sequence to attach TALE DNA Binding domain to resolvase
SEH10a-B	43 bp	AGTCAGCAGCGCCTCAAGTGCCCGGGCTCCCGACCACTGTTTA	linker sequence to attach TALE DNA Binding domain to resolvase
SEH10b-T	47 bp	CCGCGGGCCTCCGCTGCAGCTCGACACCGGGCAGCTTTTAAAGATC G	linker sequence to attach TALE DNA Binding domain to resolvase
SEH10b-B	61 bp	CGATCTTTAAAAGCTGCCCGGTGTCGAGCTGCAGCGGAGGCCCGC GGAGCTCACCCGCCAC	linker sequence to attach TALE DNA Binding domain to resolvase
SEH11-T	16 bp	CTAGTTCCGCGGGTTC	To introduce termination codon into TALER construct
SEH11-B	16 bp	AATTGAACCCGCGGAA	To introduce termination codon into TALER construct
SEH13-T	29 bp	CTAGTGCTAGCAGAAGATCTTGATCAGCG	linker sequence between TALE and resolvase, introducing restriction sites
SEH13-B	29 bp	GTACCGCTGATCAAGATCTTCTGCTAGCA	linker sequence between TALE and resolvase, introducing restriction sites
SEH14-T	13 bp	CTAGTCACGAGGC	linker sequence between TALE and resolvase
SEH14-B	13 bp	AATTGCCTCGTGA	linker sequence between TALE and resolvase

**Table 2.4:** Oligonucleotides used in this research (continued)

Oligo ID:	Size (bp)	Sequence (5'-3')	Purpose
SEH12*22-T	60 bp	CAGATGTGGAAACGGAAGAAGCAAATATTATAAATTATAGCTCTTCCGTTT CCACATCTG	To introduce TALER T-22 Site I
SEH12*22-B	68 bp	AATTCAGATGTGGAAACGGAAGAGCTATAATTTATAATATTTGCTTCTTCC GTTTCCACATCTGAGCT	To introduce TALER T-22 Site I
SEH15*18-T	56 bp	CAGATGTGGAAACGGAAGACAAATATTATAAATTATATCTTCCGTTTCCAC ATCTG	To introduce TALER T-18 Site I
SEH15*18-B	64 bp	AATTCAGATGTGGAAACGGAAGATATAATTTATAATATTTGTCTTCCGTTT CCACATCTGAGCT	To introduce TALER T-18 Site I
SEH16*20-T	58 bp	CAGATGTGGAAACGGAAGAGCAAATATTATAAATTATAGTCTTCCGTTTCC ACATCTG	To introduce TALER T-20 Site I
SEH16*20-B	66 bp	AATTCAGATGTGGAAACGGAAGACTATAATTTATAATATTTGCTCTTCCGT TTCCACATCTGAGCT	To introduce TALER T-20 Site I
SEH17*24-T	62 bp	CAGATGTGGAAACGGAAGACAGCAAATATTATAAATTATAGCGTCTTCCG TTTCCACATCTG	To introduce TALER T-24 Site I
SEH17*24-B	70 bp	AATTCAGATGTGGAAACGGAAGACGCTATAATTTATAATATTTGCTGTCTT CCGTTTCCACATCTGAGCT	To introduce TALER T-24 Site I
SEH18*26-T	64 bp	CAGATGTGGAAACGGAAGAACAGCAAATATTATAAATTATAGCGAT CTTCCGTTTCCACATCTG	To introduce TALER T-26 Site I
SEH18*26-B	72 bp	AATTCAGATGTGGAAACGGAAGATCGCTATAATTTATAATATTTGCT GTTCTTCCGTTTCCACATCTGAGCT	To introduce TALER T-26 Site I
SEH19*28-T	66 bp	CAGATGTGGAAACGGAAGAGACAGCAAATATTATAAATTATAGCGA GTCTTCCGTTTCCACATCTG	To introduce TALER T-28 Site I
SEH19*28-B	74 bp	AATTCAGATGTGGAAACGGAAGACTCGCTATAATTTATAATATTTGC TGTCTCTTCCGTTTCCACATCTGAGCT	To introduce TALER T-28 Site I
SEH25a-T	111 bp	AATTGTCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACC ATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATG CAGTCAAAAAGGGTCTGCCTC	To introduce Histidine (H <sub>6</sub> ) Tag to TALER construct
SEH25a-B	87 bp	GCATCGAGCGCGGGTCGTCCACCAAGACATGCCAACGCCACCAG ATGGTCATTCGTTAACGCAGCCAACGCGGGATCGGGCCTGGAC	To introduce Histidine (H <sub>6</sub> ) Tag to TALER construct

**Table 2.4:** Oligonucleotides used in this research (continued)

Oligo ID:	Size (bp)	Sequence (5'-3')	Purpose
SEH25b-T	90 bp	ATGCTCCCGCATTGATCAAAAGAACCAACCGGCGG ATTCCCGAGAGAACTTCCCATCGAGTCGCGCATCA CCATCACCATCACTGATCAG	To introduce Histidine (H <sub>6</sub> ) Tag to TALER construct
SEH25b-B	114 bp	GATCCTGATCAGTGATGGTGATGGTGATGCGCGACTCGATGGGA AGTTCTCTCGGGAATCCGCCGGTTGGTTCTTTTGATCAATGCGGG AGCATGAGGCAGACCCTTTTTGACT	To introduce Histidine (H <sub>6</sub> ) Tag to TALER construct
SEH27-T	9 bp	CTAGTATGT	To make TALER $\Delta$ 48 truncation
SEH27-B	10 bp	CCGGGACATA	To make TALER $\Delta$ 48 truncation
SEH28-T	25 bp	CTAGTCTTCTTGATTCGATGCCCGC	To make TALER $\Delta$ 84 truncation
SEH28-B	19 bp	GGGCATCGAATCAAGAAGA	To make TALER $\Delta$ 84 truncation
SEH29-T	11 bp	CTAGTCCCACC	To make TALER $\Delta$ 119 truncation
SEH29-B	11 bp	GTACGGTGCGGA	To make TALER $\Delta$ 119 truncation
SEH30-T	18 bp	CTAGTCCGGCAGCACAGG	To make TALER $\Delta$ 149 truncation
SEH30-B	18 bp	TCGACCTGCGCTGCCGGA	To make TALER $\Delta$ 149 truncation
SEH31-T	30 bp	TATGAAAGACCTAATCATTCGAATATTAGC	To make Sin Q115R $\Delta$ 109 tadpole
SEH31-B	32bp	CATGGCTAATATTCGAATGATTAGGTCTTTCA	To make Sin Q115R $\Delta$ 109 tadpole

**Table 2.4:** Oligonucleotides used in this research (continued)

Oligo ID:	Size (bp)	Sequence (5'-3')	Purpose
SEH32-T	30 bp	TATGGCAGACCTAATCATTCAAATATTAGC	To make Sin WT $\Delta$ 109 K110A E-helix mutant tadpole
SEH32-B	32 bp	CATGGCTAATATTTGAATGATTAGGTCTGCCA	To make Sin WT $\Delta$ 109 K110A E-helix mutant tadpole
SEH33-T	30 bp	TATGAAAGACGCAATCATTCAAATATTAGC	To make Sin WT $\Delta$ 109 L112A E-helix mutant tadpole
SEH33-B	32 bp	CATGGCTAATATTTGAATGATTGCGTCTTTCA	To make Sin WT $\Delta$ 109 L112A E-helix mutant tadpole
SEH34-T	30 bp	TATGAAAGACCTAATCGCTCAAATATTAGC	To make Sin WT $\Delta$ 109 I114A E-helix mutant tadpole
SEH34-B	32 bp	CATGGCTAATATTTGAGCGATTAGGTCTTTCA	To make Sin WT $\Delta$ 109 I114A E-helix mutant tadpole
SEH35-T	30 bp	TATGAAAGACCTAATCATTCAAGCATTAGC	To make Sin WT $\Delta$ 109 I116A E-helix mutant tadpole
SEH35-B	32 bp	CATGGCTAATGCTTGAATGATTAGGTCTTTCA	To make Sin WT $\Delta$ 109 I116A E-helix mutant tadpole
SEH36-T	66 bp	TATGAAAGACCTAATCATTCAAATATTAGCCGCGGTTTCGGAACAA GAACGAAATGAAAGTAAACG	To make Sin WT $\Delta$ 109 M119A E-helix mutant tadpole
SEH36-B	68 bp	TCGACGTTTACTTTTCATTTTCGTTCTTGTTCCGAAACCGCGGCTAAT ATTTGAATGATTAGGTCTTTCA	To make Sin WT $\Delta$ 109 M119A E-helix mutant tadpole
SEH37-T	66 bp	TATGAAAGACCTAATCATTCAAATATTAGCCATGGTTGCGGAACAA GAACGAAATGAAAGTAAACG	To make Sin WT $\Delta$ 109 S121A E-helix mutant tadpole
SEH37-B	68 bp	TCGACGTTTACTTTTCATTTTCGTTCTTGTTCCGCAACCATGGCTAAT ATTTGAATGATTAGGTCTTTCA	To make Sin WT $\Delta$ 109 S121A E-helix mutant tadpole

**Table 2.4:** Oligonucleotides used in this research (continued)

Oligo ID:	Size (bp)	Sequence (5'-3')	Purpose
SEH40-T	52 bp	GTAACCCATTACTAGATAAAATTTATGGCAGACCTAATCATTCGAAT ATTAGC	To make Full-length Sin Q115R K110A E-helix mutant
SEH40-B	51 bp	CATGGCTAATATTCGAATGATTAGGTCTGCCATAAATTTATCTAGT AATGG	To make Full-length Sin Q115R K110A E-helix mutant
SEH41-T	52 bp	GTAACCCATTACTAGATAAAATTTATGAAAGACGCAATCATTCGAAT ATTAGC	To make Full-length Sin Q115R L112A E-helix mutant
SEH41-B	51 bp	CATGGCTAATATTCGAATGATTGCGTCTTTCATAAATTTATCTAGTA ATGG	To make Full-length Sin Q115R L112A E-helix mutant
SEH42-T	52 bp	GTAACCCATTACTAGATAAAATTTATGAAAGACCTAATCGCTCGAAT ATTAGC	To make Full-length Sin Q115R I114A E-helix mutant
SEH42-B	51 bp	CATGGCTAATATTCGAGCGATTAGGTCTTTCATAAATTTATCTAGT AATGG	To make Full-length Sin Q115R I114A E-helix mutant
SEH43-T	52 bp	GTAACCCATTACTAGATAAAATTTATGAAAGACCTAATCATTCGAGC ATTAGC	To make Full-length Sin Q115R I116A E-helix mutant
SEH43-B	51 bp	CATGGCTAATGCTCGAATGATTAGGTCTTTCATAAATTTATCTAGT AATGG	To make Full-length Sin Q115R I116A E-helix mutant
SEH44-T	88 bp	GTAACCCATTACTAGATAAAATTTATGAAAGACCTAATCATTCGAATA TTAGCCGCGGTTTCGGAACAAGAACGAAATGAAAGTAAACG	To make Full-length Sin Q115R M119A E-helix mutant
SEH44-B	87 bp	TCGACGTTTACTTTTCATTTTCGTTCTTGTTCCGAAACCGCGGCTAAT ATTCGAATGATTAGGTCTTTCATAAATTTATCTAGTAATGG	To make Full-length Sin Q115R M119A E-helix mutant
SEH45-T	36 pb	CATGGTTGCGGAACAAGAACGAAATGAAAGTAAACG	To make Full-length Sin Q115R S121A E-helix mutant
SEH45-B	36 bp	TCGACGTTTACTTTTCATTTTCGTTCTTGTTCCGCAAC	To make Full-length Sin Q115R S121A E-helix mutant

**Table 2.4:** Oligonucleotides used in this research (continued)

Oligo ID:	Size (bp)	Sequence (5'-3')	Purpose
SEH48-T	75 bp	CATGCCGGGAAGATGTGGAAACGGAAGAAGCAAATATTATAAATT ATAGCTCTTCCGTTTCCACATCTAAGAATT	T-22 TALER Site I to be 5' end labelled for In vitro binding assays
SEH48-B	76 bp	CTAGAATTCTTAGATGTGGAAACGGAAGAGCTATAATTTATAATAT TTGCTTCTTCCGTTTCCACATCTCTCCCGG	T-22 TALER Site I to be 5' end labelled for In vitro binding assays
SEH49-T	58 bp	CATGCCGGGAGCTCACTAGTATGTATGATTAGGGTATACCCTAAT CATACAAAGAATT	58bp RSinR Site I to be 5' end labelled for In vitro binding assays
SEH49-B	58 bp	CTAGAATTCTTTGTATGATTAGGGTATACCCTAATCATACATACTA GTGAGCTCCCGG	58bp RSinR Site I to be 5' end labelled for In vitro binding assays
SEH50-T	58 bp	CATGCCGGGAGCTCACTAGTATGTATGATTATATTATAATATAATC ATACAAAGAATT	RTn3R Site I to be 5' end labelled for In vitro binding assays
SEH50-B	58 bp	CTAGAATTCTTTGTATGATTATATTATAATATAATCATACATACTAGT GAGCTCCCGG	RTn3R Site I to be 5' end labelled for In vitro binding assays
SEH51-T	58 bp	CATGCCGGGAGCTCACTAGTAAACGGCGGCAAATCGTACCCTAAT CATACAAAGAATT	$\Delta$ R Site I to be 5' end labelled for In vitro binding assays
SEH51-B	58 bp	CTAGAATTCTTTGTATGATTAGGGTACGATTTGCCGCCGTTTACTA GTGAGCTCCCGG	$\Delta$ R Site I to be 5' end labelled for In vitro binding assays
SEH52-T	45 bp	CATGCCGGGATGTATGATTAGGGTATACCCTAATCATACAGAATT	45bp RSinR Site I to be 5' end labelled for In vitro binding assays
SEH52-B	45 bp	CTAGAATTCTGTATGATTAGGGTATACCCTAATCATACATCCCGG	45bp RSinR Site I to be 5' end labelled for In vitro binding assays
SEH53-T	77 bp	CATGCCGCTAGCCGCCTAGGCTATCAGTATGTATGATTAGGGTAT ACCCTAATCATACAAAGAGCGTCGCTACTATT	77bp RSinR Site I to be 5' end labelled for In vitro binding assays
SEH53-B	77 bp	CTAGAATAGTAGCGACGCTCTTTGTATGATTAGGGTATACCCTAAT CATACATACTGATAGCCTAGGCGGCTAGCGG	77bp RSinR Site I to be 5' end labelled for In vitro binding assays

**Table 2.4:** Oligonucleotides used in this research (continued)



Plasmid	Antibiotic Marker	Size (bp)	Description	Source
pCO1	Ap <sup>R</sup>	2543	pMTL23 with Tn3 Site I cloned in between EcoRI and SacI	C.Muir/M.Stark
pDB34	Km <sup>R</sup>	8371	<i>In vivo</i> recombination substrate, containing two copies of Tn3 <i>res</i> site in direct repeat, flanking a GalK gene	Arnold et al.
pMS140	Ap <sup>R</sup>	5469	pAT5Δ derivative encoding the Tn3 resolvase ORF, with low level expression suitable for <i>in vivo</i> recombination assays	M.Stark
pMS183Δ	Km <sup>R</sup>	4863	<i>In vivo</i> recombination substrate precursor plasmid, allows cloning of recombination between BglII/BsrGI and NcoI/XbaI flanking galK indicator	M.Prorocic
pMTL23	Ap <sup>R</sup>	2505	Cloning vector	Chambers et al. 1988
pSA1101	Km <sup>R</sup>	6692	Wild type Tn3 resolvase expression plasmid with inducible T7 promoter	Arnold et al. 1999
pSA1121	Km <sup>R</sup>	6711	Wild type Tn3 resolvase expression plasmid with inducible T7 promoter and C-terminal Histidine tag	Rowland et al. 2005
pSA1171	Km <sup>R</sup>	6762	Q115R Sin resolvase expression plasmid with inducible T7 promoter and C-terminal Histidine tag	S.Rowland
pSA1195	Km <sup>R</sup>	6438	Δ109 [V138A] Sin resolvase expression plasmid with inducible T7 promoter and C-terminal Histidine tag	S.Rowland
pSA1196	Km <sup>R</sup>	6393	Δ124 [V138A] Sin resolvase expression plasmid with inducible T7 promoter and C-terminal Histidine tag	S.Rowland
pSA9944	Ap <sup>R</sup>	5535	Low level expression plasmid, encoding wild type Sin resolvase	Mouw et al. 2008

**Table 2.5:** Plasmids constructed for this research and their precursors. Plasmid names are highlighted according to plasmid type: Yellow = pMTL23-based single site recombination plasmids, light green = pMS140 based low level resolvase expression plasmids, dark green = pSA1101-derived resolvase overexpression plasmids, blue = 2-site *in vitro* recombination substrate plasmid, red = 2-site *in vivo* recombination substrate plasmid, purple = high copy number 2-site *in vitro* substrate.

Plasmid	Antibiotic Marker	Size (bp)	Description	Source
pSA9994	Ap <sup>R</sup>	5539	Low level expression plasmid encoding Sin resolvase Q115R mutant	S.Rowland
pSE1005	Ap <sup>R</sup>	2544	pMTL23 with oligo for symmetrised Sin Site I (RSinR) cloned in between AlwNI-SacI	Section2.7.iv
pSE1006	Ap <sup>R</sup>	2544	pMTL23 with oligo for symmetrised Tn3 Site I (RTn3R) cloned in between AlwNI-SacI	Section2.7.iv
pSE1008	Ap <sup>R</sup>	2544	pMTL23 with oligo for Sin Half-Site I ( $\Delta$ R) cloned in between AlwNI-SacI	Section2.7.iv
pSE1012	Km <sup>R</sup>	4950	<i>In vivo</i> recombination substrate, containing two copies of symmetrised Sin Site I in direct repeat, flanking a GalK gene	Section 2.7.v
pSE1013	Km <sup>R</sup>	4950	<i>In vivo</i> recombination substrate, containing two copies of symmetrised Tn3 Site I in direct repeat, flanking a GalK gene	Section 2.7.v
pSE1014	Km <sup>R</sup>	4950	<i>In vivo</i> recombination substrate, containing two copies of Sin Half-Site I in direct repeat, flanking a GalK gene	Section 2.7.v
pSE1020	Ap <sup>R</sup>	2563	pMTL23 with oligo for T-22 NM Site I cloned in between EcoRI-SacI	Section2.7.iv
pSE1021	Km <sup>R</sup>	4984	<i>In vivo</i> recombination substrate, containing two copies of T-22 NM Site I in direct repeat, flanking a GalK gene	Section 2.7.v
pSE1022	Ap <sup>R</sup>	2559	pMTL23 with oligo for T-18 NM Site I cloned in between EcoRI-SacI	Section2.7.iv
pSE1023	Ap <sup>R</sup>	2561	pMTL23 with oligo for T-20 NM Site I cloned in between EcoRI-SacI	Section2.7.iv

**Table 2.5** (continued).

Plasmid	Antibiotic Marker	Size (bp)	Description	Source
pSE1024	Ap <sup>R</sup>	2565	pMTL23 with oligo for T-24 NM Site I cloned in between EcoRI-SacI	Section2.7.iv
pSE1025	Ap <sup>R</sup>	2567	pMTL23 with oligo for T-26 NM Site I cloned in between EcoRI-SacI	Section2.7.iv
pSE1026	Ap <sup>R</sup>	2569	pMTL23 with oligo for T-28 NM Site I cloned in between EcoRI-SacI	Section2.7.iv
pSE1027	Km <sup>R</sup>	4976	<i>In vivo</i> recombination substrate, containing two copies of T-18 NM Site I in direct repeat, flanking a GalK gene	Section 2.7.v
pSE1028	Km <sup>R</sup>	4980	<i>In vivo</i> recombination substrate, containing two copies of T-20 NM Site I in direct repeat, flanking a GalK gene	Section 2.7.v
pSE1029	Km <sup>R</sup>	4988	<i>In vivo</i> recombination substrate, containing two copies of T-24 NM Site I in direct repeat, flanking a GalK gene	Section 2.7.v
pSE1030	Km <sup>R</sup>	4992	<i>In vivo</i> recombination substrate, containing two copies of T-26 NM Site I in direct repeat, flanking a GalK gene	Section 2.7.v
pSE1031	Km <sup>R</sup>	4996	<i>In vivo</i> recombination substrate, containing two copies of T-28 NM Site I in direct repeat, flanking a GalK gene	Section 2.7.v
pSE1032	Km <sup>R</sup>	4943	pSB424 based <i>in vivo</i> binding substrate with one copy of T-24 NM site I containing GalK promoter	Section2.7.iv
pSE1033	Km <sup>R</sup>	4945	pSB424 based <i>in vivo</i> binding substrate with one copy of T-26 NM site I containing GalK promoter	Section2.7.iv
pSE1034	Km <sup>R</sup>	4947	pSB424 based <i>in vivo</i> binding substrate with one copy of T-28 NM site I containing GalK promoter	Section2.7.iv

**Table 2.5** (continued).

Plasmid	Antibiotic Marker	Size (bp)	Description	Source
pUC4K	Kan <sup>R</sup> /Amp <sup>R</sup>	3914	High copy number cloning vector used for construction of <i>in vitro</i> substrates	Amersham bioscience
pSE1041	Kan <sup>R</sup> /Amp <sup>R</sup>	3955	High copy number <i>in vitro</i> substrate with two copies of T-20 site I	Section 4.4
pSE1042	Kan <sup>R</sup> /Amp <sup>R</sup>	3959	High copy number <i>in vitro</i> substrate with two copies of T-22 site I	Section 4.4
pSE1043	Kan <sup>R</sup> /Amp <sup>R</sup>	3963	High copy number <i>in vitro</i> substrate with two copies of T-24 site I	Section 4.4
pSE1044	Kan <sup>R</sup> /Amp <sup>R</sup>	3967	High copy number <i>in vitro</i> substrate with two copies of T-26 site I	Section 4.4
pSE1045	Kan <sup>R</sup> /Amp <sup>R</sup>	3971	High copy number <i>in vitro</i> substrate with two copies of T-28 site I	Section 4.4
pSE2001	Km <sup>R</sup>	6462	Truncated ( $\Delta$ M101) Sin resolvase expression plasmid with inducible T7 promoter	Section 3.2.ii
pSE2003	Km <sup>R</sup>	6438	Truncated ( $\Delta$ M109) Sin resolvase expression plasmid with inducible T7 promoter	Section 3.2.ii
pSE2004	Km <sup>R</sup>	6393	Truncated ( $\Delta$ M124) Sin resolvase expression plasmid with inducible T7 promoter	Section 3.2.ii
pSE2005	Km <sup>R</sup>	6405	Wild Type Tn3 resolvase expression plasmid with inducible T7 promoter	Section 3.2.ii
pSE2028	Ap <sup>R</sup>	5636	Low copy number expression plasmid, pMS140 with pAMC1-4 (Tn3 ZFR) cloned in between EcoRI and Asp718	Section 2.7.iii

**Table 2.5** (continued).

Plasmid	Antibiotic Marker	Size (bp)	Description	Source
pSE2029	Ap <sup>R</sup>	5382	Low copy number expression plasmid, pSE2028 with oligo to introduce restriction sites for linker insertion	Section 2.7.iii
pSE2031	Ap <sup>R</sup>	7452	Low copy number expression plasmid with TAL 1297 truncated at position 222	Section 2.7.iii
pSE2033	Km <sup>R</sup>	8675	TALER TAL222 expression plasmid with inducible T7 promoter	Section 2.7.i
pSE2034	Ap <sup>R</sup>	7640	Low copy number TALER TAL-222 expression plasmid with a 6aa GS linker	Section 2.7.iii
pSE2035	Ap <sup>R</sup>	7644	Low copy number TALER TAL-222 expression plasmid with a 9aa GS linker	Section 2.7.iii
pSE2036	Ap <sup>R</sup>	7648	Low copy number TALER TAL-222 expression plasmid with a 12aa GS linker	Section 2.7.iii
pSE2037	Ap <sup>R</sup>	7652	Low copy number TALER TAL-222 expression plasmid with a 7aa GS linker	Section 2.7.iii
pSE2040	Ap <sup>R</sup>	7656	Low copy number TALER Δ154 expression plasmid, made from pSE2030 with oligo SEH20 cloned in between SpeI-Bsu36I	Section 2.7.iii
pSE2041	Ap <sup>R</sup>	7662	TALER Δ154 low copy number expression plasmid with a 6aa GS linker	Section 2.7.iii
pSE2050	Km <sup>R</sup>	8879	TALER Δ154 expression plasmid with inducible T7 promoter	Section 2.7.i
pSE2051	Km <sup>R</sup>	8885	TALER Δ154 expression plasmid with a 6aa GS linker and an inducible T7 promoter	Section 2.7.i
pSE2060	Km <sup>R</sup>	9354	Tn3-TALER with 6aa linker and GeneArt TAL 1-(+63)H6 with inducible T7 promoter	Section 2.7.i

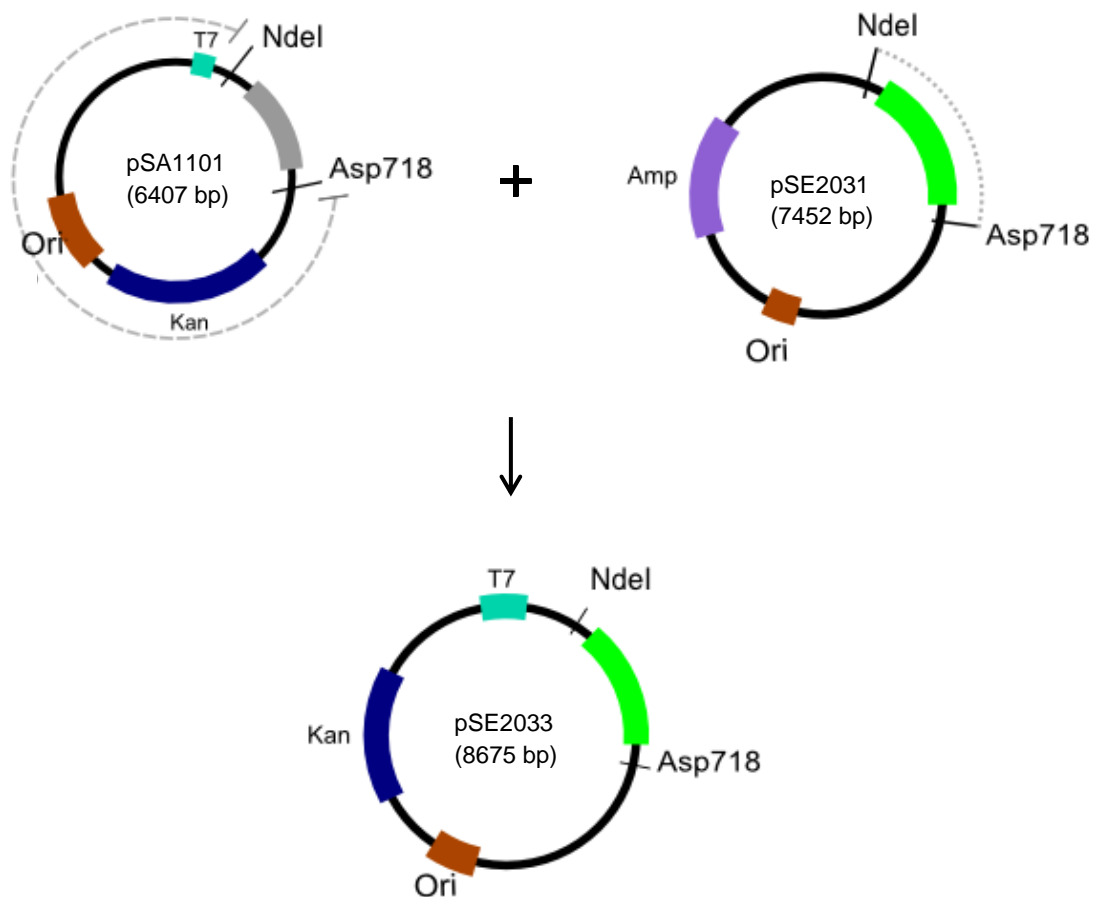
**Table 2.5** (continued).

Plasmid	Antibiotic Marker	Size (bp)	Description	Source
pSE2065	Ap <sup>R</sup>	8131	Low copy number plasmid expressing Tn3-TALER with 6aa linker and GeneArt TAL 1-(+63)H6	Section 2.7.iii
pSE2080	Km <sup>R</sup>	9213	NM resolvase 1-144 with 6aa linker and GeneArt TAL $\Delta$ 48-(+63)H6 with T7 inducible promoter	Section 2.7.i
pSE2085	Ap <sup>R</sup>	7990	Low copy number plasmid expressing NM resolvase 1-144 with 6aa linker and GeneArt TAL $\Delta$ 48-(+63)H6	Section 2.7.iii
pSE2090	Km <sup>R</sup>	9105	NM resolvase 1-144 with 6aa linker and GeneArt TAL $\Delta$ 84-(+63)H6 with T7 inducible promoter	Section 2.7.i
pSE2095	Ap <sup>R</sup>	7882	Low copy number plasmid expressing NM resolvase 1-144 with 6aa linker and GeneArt TAL $\Delta$ 84-(+63)H6	Section 2.7.iii
pSE2100	Km <sup>R</sup>	9003	NM resolvase 1-144 with 6aa linker and GeneArt TAL $\Delta$ 119-(+63)H6 with T7 inducible promoter	Section 2.7.i
pSE2105	Ap <sup>R</sup>	7777	Low copy number plasmid expressing NM resolvase 1-144 with 6aa linker and GeneArt TAL $\Delta$ 119-(+63)H6	Section 2.7.iii
pSE2110	Km <sup>R</sup>	8913	NM resolvase 1-144 with 6aa linker and GeneArt TAL $\Delta$ 149-(+63)H6 with T7 inducible promoter	Section 2.7.i
pSE2115	Ap <sup>R</sup>	7687	Low copy number plasmid expressing NM resolvase 1-144 with 6aa linker and GeneArt TAL $\Delta$ 149-(+63)H6	Section 2.7.iii
pSE2300	Km <sup>R</sup>	6438	Truncated ( $\Delta$ M109) His-6 tagged Q115R Sin resolvase expression plasmid with inducible T7 promoter	Section 3.3.i
pSE2301	Km <sup>R</sup>	6438	Truncated ( $\Delta$ M109) His-6 tagged Sin resolvase expression plasmid with K110A E-helix mutation and inducible T7 promoter	Section 3.3.i

**Table 2.5** (continued).

Plasmid	Antibiotic Marker	Size (bp)	Description	Source
pSE2302	Km <sup>R</sup>	6438	Truncated ( $\Delta$ M109) His-6 tagged Sin resolvase expression plasmid with L112A E-helix mutation and inducible T7 promoter	Section3.3.i
pSE2303	Km <sup>R</sup>	6438	Truncated ( $\Delta$ M109) His-6 tagged Sin resolvase expression plasmid with I114A E-helix mutation and inducible T7 promoter	Section3.3.i
pSE2304	Km <sup>R</sup>	6438	Truncated ( $\Delta$ M109) His-6 tagged Sin resolvase expression plasmid with I116A E-helix mutation and inducible T7 promoter	Section3.3.i
pSE2305	Km <sup>R</sup>	6438	Truncated ( $\Delta$ M109) His-6 tagged Sin resolvase expression plasmid with M119A E-helix mutation and inducible T7 promoter	Section3.3.i
pSE2306	Km <sup>R</sup>	6438	Truncated ( $\Delta$ M109) His-6 tagged Sin resolvase expression plasmid with S121A E-helix mutation and inducible T7 promoter	Section3.3.i
pSE2315	Km <sup>R</sup>	6762	His-6 tagged Q115R Sin resolvase expression plasmid with inducible T7 promoter	Section3.3.i
pSE2316	Km <sup>R</sup>	6762	His-6 tagged Q115R Sin resolvase expression plasmid with K110A E-helix mutation and inducible T7 promoter inducible T7 promoter	Section3.3.i
pSE2317	Km <sup>R</sup>	6762	His-6 tagged Q115R Sin resolvase expression plasmid with L112A E-helix mutation and inducible T7 promoter	Section3.3.i
pSE2318	Km <sup>R</sup>	6762	His-6 tagged Q115R Sin resolvase expression plasmid with I114A E-helix mutation and inducible T7 promoter	Section3.3.i
pSE2319	Km <sup>R</sup>	6762	His-6 tagged Q115R Sin resolvase expression plasmid with I116A E-helix mutation and inducible T7 promoter	Section3.3.i
pSE2320	Km <sup>R</sup>	6762	His-6 tagged Q115R Sin resolvase expression plasmid with M119A E-helix mutation and inducible T7 promoter	Section3.3.i
pSE2321	Km <sup>R</sup>	6762	His-6 tagged Q115R Sin resolvase expression plasmid with S121A E-helix mutation and inducible T7 promoter	Section3.3.i

**Table 2.5** (Continued).



**Figure 2.1:** High level expression plasmids constructed based on pSA1101. High-level expression plasmids have a ColEI origin of replication, a kanamycin resistance marker and an IPTG-inducible T7 promoter. A resolvase open reading frame (green) is cloned in between NdeI and Asp718 in the pSA1101 backbone.



### **2.7.ii Low level resolvase expression plasmids**

All low level expression plasmids were constructed based on a pMS140 backbone (Burke et al 2004). pMS140 has a ColEI origin of replication, contains an ampicillin resistance marker and expresses a cassetted version of wild-type Tn3 resolvase, cloned between the restriction sites NdeI and Asp718. Expression of resolvase in pMS140-derived plasmids is driven by an unidentified promoter situated within 400 bp upstream of the NdeI site. The low level of expression is suitable for *in vivo* recombination assays. pMS140 and all of its derivatives contain a unique EagI restriction site within the resolvase reading frame after the conserved GR motif (amino acid residues G141 and R142). This allows the swapping of the C-terminus of Tn3 resolvase with DNA-binding domains from other proteins, such as Sin in Tn3/Sin hybrid proteins or TALE domains in TALER hybrids.

### **2.7.iii *In vitro* recombination substrate plasmids**

pMTL23 is the precursor to the *in vivo* and *in vitro* two-site plasmids. pMTL23 is a high copy number cloning vector containing an ampicillin resistance marker and a deregulated ColEI origin of replication (Chambers et al. 1988). It contains a large polylinker, with many useful restriction sites. The *res* sites utilised in this study were cloned into the pMTL23 vector as oligonucleotides (MB67/71, MB68/72 and MB70/74; see table 2.4) between the restriction sites AlwNI-SacI. The *res* site can be excised from this plasmid between the restriction sites XbaI-NcoI and BamHI-Asp718 for construction of 2-site substrates for *in vivo* assays (section 2.7.iv).

## **2.7.iv *In vivo* recombination substrate plasmids**

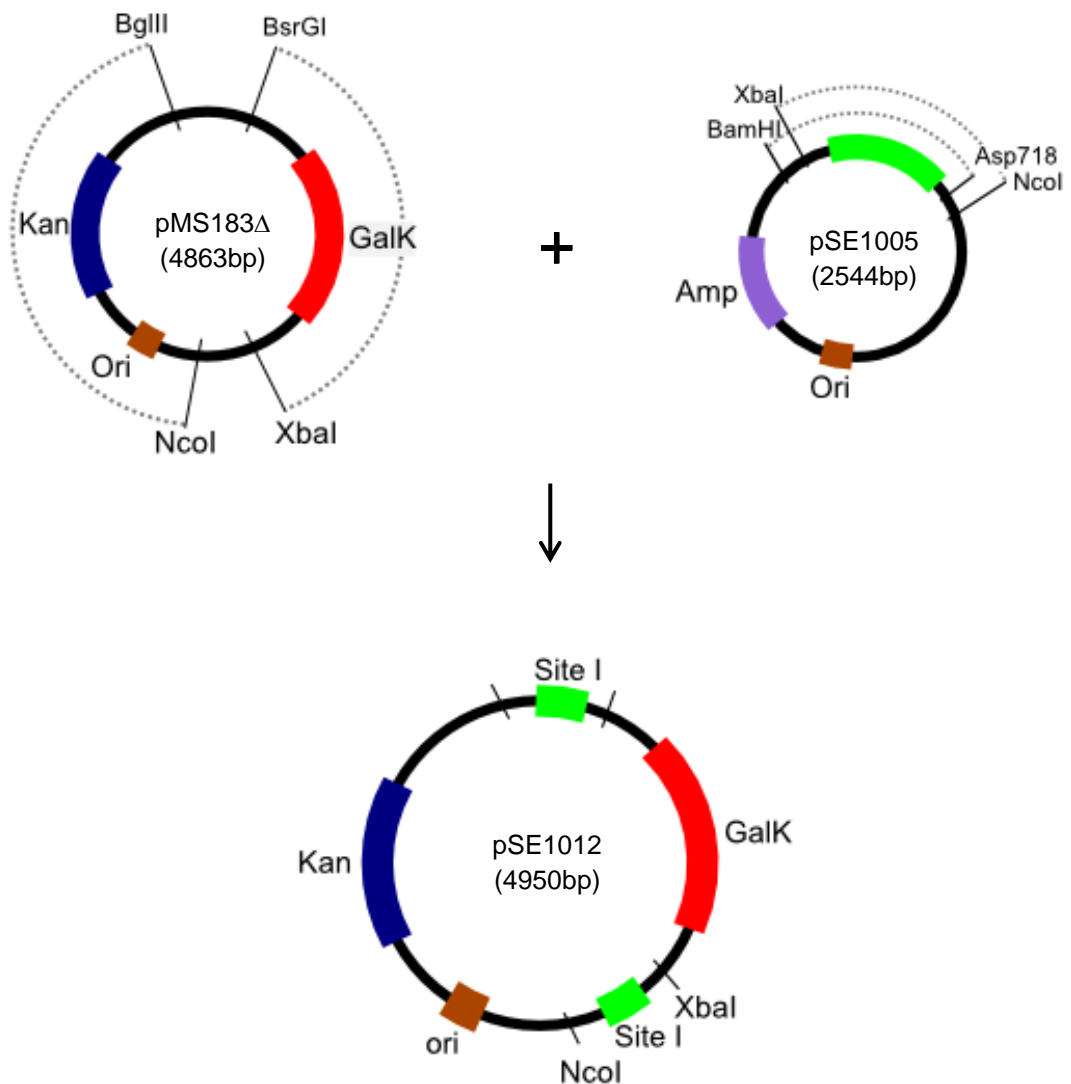
All substrate plasmids used for *in vivo* assays in this study are based on a pMS183 $\Delta$  backbone. pMS183 $\Delta$  expresses a kanamycin resistance marker and has a pSC101 replication origin. *In vivo* recombination substrates were constructed by a four-piece ligation, with BglII-NcoI and BsrGI-XbaI from pMS183 $\Delta$  and XbaI-NcoI and BamHI-Asp718 from a pMTL23-based single site plasmid (Figure 2.2). This cloning is made possible by the compatibility of the sticky ends made by the enzymes BsrGI and Asp718, and BglII and BamHI respectively.

## **2.8 Making *E.coli* cells competent**

Competent *E.coli* cells for DNA transformation were prepared routinely by chemical treatment of log phase cultures (section 2.8.i). When higher transformation efficiency was required, cells were prepared by an alternate method making them suitable for electroporation (section 2.8.ii).

### **2.8.i Chemically competent cells**

Chemically competent cells were prepared by inoculating 10 ml of L-broth with 200  $\mu$ l of an overnight *E.coli* culture and growing at 37°C with shaking at 250 rpm (New Brunswick Scientific Excella E24 Incubator Shaker) to an OD<sub>600</sub> of 0.4-0.5 (mid-log phase). The culture was divided into four aliquots of 1.2 ml and chilled on ice. Cells were harvested by centrifugation in a pre-cooled rotor at 4°C for 1 minute at 9300 x g, after which the supernatant was removed. The cell pellet was re-suspended in 1.2 ml ice-cold 50 mM CaCl<sub>2</sub> and the centrifugation step was repeated. Following the second centrifugation step, the supernatant was removed and the cell pellet was re-suspended in 200  $\mu$ l ice-cold 50 mM CaCl<sub>2</sub>. Cells prepared by this method were stored on ice until use, and remained competent for up to 48 hours.



**Figure 2.2:** *In vivo* 2-site substrate construction.

*res* sites from a pMTL23-based plasmid were cloned into the *galK* expressing plasmid pMS183Δ on BamHI-Asp718 and XbaI-NcoI fragments. This cloning was made possible by the compatibility of the enzymes BsrGI and Asp718 and of BglII and BamHI.

## **2.8.ii Electro-competent cells**

To prepare electro-competent cells 400 ml of L-broth was inoculated with 2ml of an overnight culture, cells were then grown at 37°C with shaking at 250 rpm (New Brunswick Scientific Excella E24 Incubator Shaker) to an OD<sub>600</sub> of 0.4-0.5 (mid-log phase). The culture was then divided into two 200 ml aliquots and chilled on ice. Cells were harvested by centrifugation in a pre-cooled rotor (Beckman Coulter JA-14) at 4°C for 15 mins at 4000 x g. The supernatant was removed, the cell pellet was re-suspended in 400 ml ice-cold 10% glycerol and centrifugation was repeated. The resulting cell pellet was re-suspended in 200 ml ice cold 10% glycerol and the centrifugation repeated. Following this, the cell pellet was re-suspended in 20 ml ice-cold 10% glycerol, and the suspension was transferred to a 30 ml polypropylene tube and centrifuged (Beckman coulter JA-20) at 4°C for 15 mins at 4000 rpm. Finally, the pellet was re-suspended in 2 ml ice-cold 10% glycerol. The cells were divided into 20 µl aliquots, flash-frozen in liquid nitrogen and stored at 70 °C. Cells thus prepared can be stored for up to six months.

## **2.9 Transformation of competent cells**

### **2.9.i Transformation of chemically competent *E.coli* cells**

100 µl of chemically competent cells was mixed with 0.01-0.1 µg of plasmid DNA and the mixture was incubated on ice for 20 mins. The cells were then heat-shocked at 42°C in a water bath for 2 mins, followed by incubation on ice for 5 mins. 1 ml of L-broth was added and the cells were incubated at 37°C for 60-90 mins to allow expression of the plasmid. Aliquots of this liquid culture were spread on selective L-agar plates (section 2.2) and incubated at 37°C overnight.

## **2.9.ii Transformation of electro-competent cells**

0.01 µg of DNA was added to a 20 µl aliquot of cells, and transferred to an ice-cold electroporation cuvette. An electrical pulse was delivered by a Biorad Micropulser and 1 ml L-broth was immediately added. The culture was then transferred to a 2 ml Nunc tube and incubated at 37°C with shaking at 250 rpm (New Brunswick Scientific Excella E24 Incubator Shaker) for 60 mins. After this recovery stage, aliquots of the culture were spread on selective agar plates and incubated overnight at 37°C.

## **2.10 Plasmid DNA preparation**

Plasmid DNA for use in most experiments was prepared on a small scale. For *in vitro* experiments, where larger quantities of DNA were required, a larger scale preparation method was used (Section 2.10.ii). Small scale DNA preparation (section 2.10.i) was carried out using a Qiagen mini-prep kit, which uses a silica-gel membrane system to bind DNA. The larger scale preparation (section 2.10.ii) used a Qiagen midi-prep kit, which uses a silica-gel beaded column system to bind the DNA.

### **2.10.i Small Scale DNA preparation**

Plasmid mini-preps were carried out using a Qiagen plasmid mini-prep kit. Each mini-prep was carried out using a pellet from 4.5 ml of culture. Other steps were carried out according to the manufacturer's instructions.

## **2.10.ii Medium Scale DNA preparation**

Plasmid midi-preps were carried out using a Qiagen midi-prep kit and a pellet from 25 ml of cell culture. The DNA preparation was carried out according to the manufacturer's instructions.

## **2.11 Ethanol precipitation of DNA**

DNA obtained by the Qiagen DNA preparation method can be ethanol-precipitated to further purify it. Ammonium acetate solution was added to a final concentration of 0.5 M, and 2.2x sample volume of 100% ethanol was then added. The mix was then incubated on ice for 15-30 minutes, followed by centrifugation at 4°C for 30 minutes at 9300 x g. The supernatant was removed and the pellet was washed with 100 µl 80% ethanol. Centrifugation was repeated at 4°C, 9300 x g for 20 minutes, the supernatant was removed and the resulting pellet was re-suspended in 20 µl TE buffer (10 mM Tris/HCl (pH8), 0.1 mM EDTA).

## **2.12 Digestion of DNA by restriction endonucleases**

Restriction endonucleases were used to digest the DNA, using 2-10 units of enzyme per microgram of DNA and utilising the supplier's recommended buffer. Digests were typically incubated at 37°C for 1 hour (unless otherwise stated in the manufacturer's instructions). SDS loading buffer (1% SDS, 50% glycerol, 0.01% bromophenol blue) was added at a 1:5 ratio following incubation to stop the reaction.

## **2.13 Gel electrophoresis**

DNA fragments were separated according to size using either agarose (section 2.13.i) or polyacrylamide (section 2.13.ii) gel electrophoresis for large supercoiled/linear fragments >200 bp in size or small (<200 bp) fragments respectively. Oligonucleotide purification and protein separation were also carried out using polyacrylamide gels (see sections 2.19 and 2.23).

### **2.13.i Agarose gel electrophoresis**

Agarose gel electrophoresis was routinely used to separate restriction digest fragments between 200 bp and 4000 bp in size. Agarose gels of 1-1.2% w/v were made by dissolving agarose (Biorad ultrapure) in 100 ml TAE buffer (40 mM Tris/ acetate, 1 mM EDTA, pH 8.5) by heating in a microwave oven. The agarose was then cooled and poured onto a gel plate and fitted with a well-forming comb. Once set, the comb was removed, samples were added to the wells and the gels were run at room temperature for 60-90 minutes at 150 V in TAE gel running buffer. The gel was then stained and visualised (section 2.16).

### **2.13.ii Polyacrylamide gel electrophoresis**

Polyacrylamide gel electrophoresis was used to separate DNA fragments <200 bp in size. Gels for separation of double stranded DNA were typically 12% acrylamide and were prepared by making 30 ml of gel solution (12% acrylamide/bis acrylamide (19:1), 3 ml 10x TBE buffer (89 mM Tris base, 2 mM EDTA, 89 mM boric acid), 15 ml H<sub>2</sub>O, 75 µl 20% ammonium persulphate (APS) and 10 µl TEMED (N,N,N',N'-tetramethylethylenediamine). For single stranded oligonucleotide purification gels, the acrylamide was typically as above with the addition of 7 M urea.

For protein separation (section 2.23), gels were routinely made at 15% (15% acrylamide/bisacrylamide (37.5:1), 375 mM Tris/HCl (pH 8.8), 0.1% SDS, 0.1% APS, 0.2% TEMED).

For binding assay gels (section 2.29), gels were typically made at 5% acrylamide, with the addition of glycerol and were prepared by making 30 ml of gel solution (5% acrylamide/bis acrylamide (19:1), 3 ml 10x TBE buffer (89 mM Tris base, 2 mM EDTA, 89 mM boric acid), 3ml glycerol, 12ml H<sub>2</sub>O, 75 µl 20% ammonium persulphate (APS) and 10 µl TEMED (N,N,N',N'-tetramethylethylenediamine)).

The gel mix was poured between sealed and clamped glass plates with 0.75 mm spacers and a well-forming comb inserted. Polymerisation was allowed to continue for 1 hour. Following polymerisation, the clamps, comb and sealing were removed and the gel was inserted into a vertical electrophoresis kit. The gel was run at 220 V for 2.5-3 h at room temperature in 1x TBE running buffer. DNA bands separated by this method were visualised by ethidium bromide staining (section 2.16).

## **2.14 DNA size marker ladder**

For estimation of DNA fragment size, a NEB 1kb or 1kb<sup>+</sup> marker ladder was used as appropriate.

## **2.15 Sample loading buffers**

Prior to loading of restriction digest samples onto an agarose gel, one volume of SDS loading buffer was added to 5 volumes of sample. For *in vitro* cleavage and resolution reactions, the same ratio of SDS-K loading buffer (1% SDS, 50% glycerol, 0.01% bromophenol blue, 1 mg/ml protease K) was



added to the sample, with the protease acting to digest the resolvase, allowing the bands to be clearly visualised.

## **2.16 Visualisation of DNA by ethidium bromide staining**

DNA bands on both agarose and polyacrylamide gels were visualised by staining with a 0.6 µg/ml ethidium bromide (EtBr) solution (15 mg/ml EtBr stock, added to 1x gel running buffer) for 30-60 minutes. Gels stained in this way were de-stained by soaking in de-ionised water for a further 30-60 minutes to remove background ethidium bromide fluorescence. EtBr-stained bands were visualised on a long wavelength (365 nm) UV transilluminator for preparative gels where bands were to be extracted for DNA purification purposes or on a short wavelength (254 nm) UV transilluminator for visualisation and photographing of gels.

## **2.17 DNA extraction from gels**

DNA was extracted from agarose gels by cutting out a gel slice, visualised by 350 nm transillumination of an ethidium bromide-stained gel using a scalpel, and filtering the DNA through a 0.45 µM Costar column by centrifugation for 1 minute at 10,000 rpm

## **2.18 Purification and annealing of oligonucleotides**

Oligonucleotides obtained from MWG Eurofins Operon were re-suspended in an appropriate volume of TE buffer (10 mM Tris/HCl, 0.1 mM EDTA) according to the provided oligonucleotide synthesis sheet, to give a final concentration of 100 µM. 1 volume of loading buffer was added to 5 volumes of oligonucleotide. Typically 30µl of 100 µM oligonucleotide/loading buffer

was loaded onto a 12% urea/acrylamide gel (section 2.13) and run in TBE buffer at 20 W (450-500 V) for 3 hours. Gels were then stained using StainsAll (1-Ethyl-2-[3-(1-ethylnaphtho[1,2-d]thiazolin-2-ylidene)-2-methylpropenyl] naphtho[1,2-d]thiazolium bromide) stain (Sigma Aldrich) and the DNA was excised in a gel slice using a scalpel. Crush and soak followed by ethanol precipitation was then used for purification of the DNA (sections 2.26 and 2.11). Purified single stranded oligonucleotides were then annealed by mixing the top and bottom strands in a 1:1 ratio with 100 mM NaCl and 1x TE buffer, heating to 9 °C.

## **2.19 Ligation of DNA fragments from restriction digests**

For cloning, DNA fragments were ligated in Invitrogen ligation buffer (50 mM Tris/HCl (pH 7.6), 10 mM MgCl<sub>2</sub>, 1 mM dithiothreitol (DTT), 5% w/v polyethylene glycol-8000) using 1 unit of T4 DNA ligase in a final volume of 20 µl. The ratio of vector to insert was typically 1:3. Ligation reactions were carried out at room temperature overnight and used to transform competent *E.coli* cells (section 2.8).

## **2.20 Sequencing plasmid DNA**

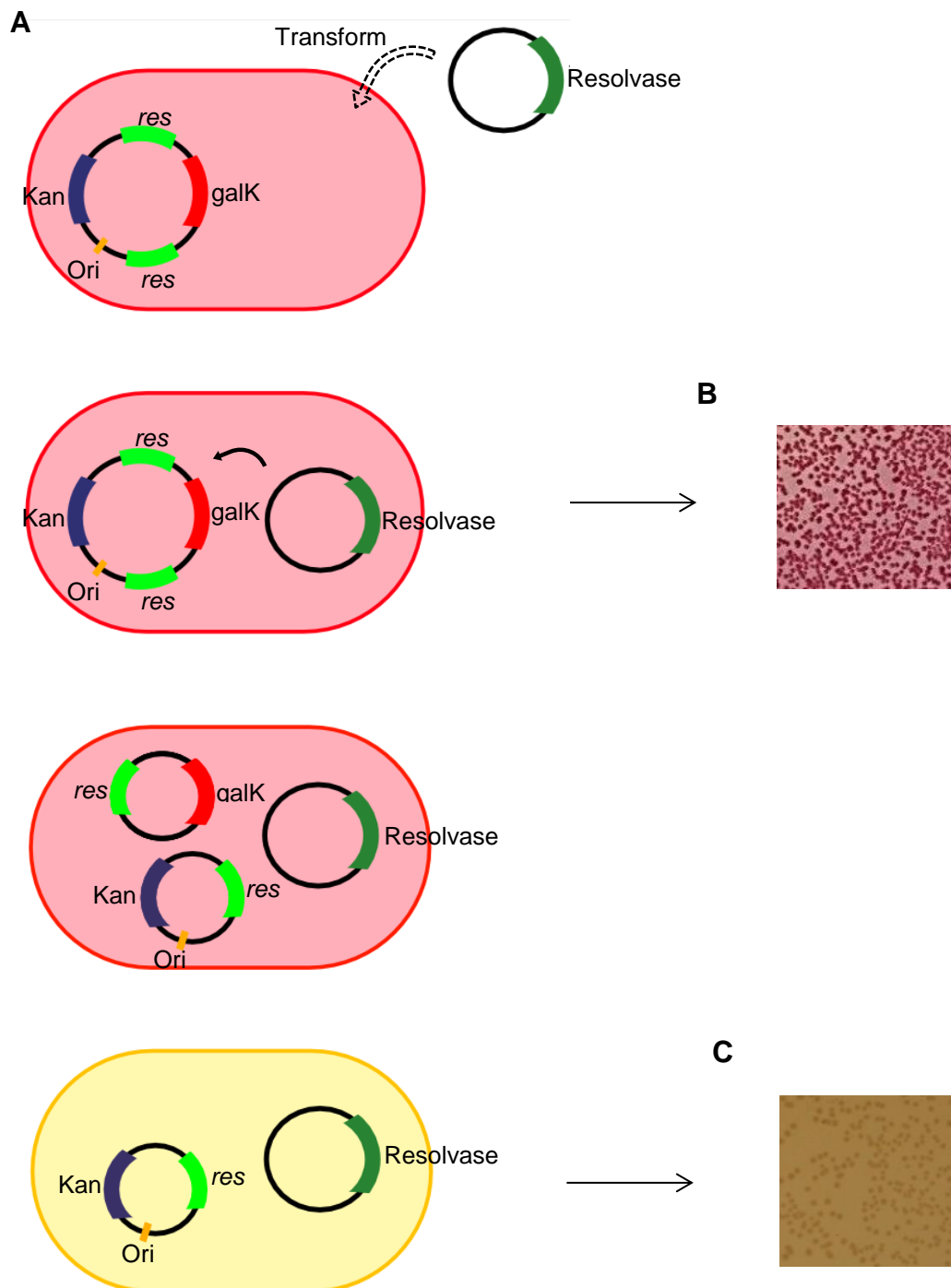
Sequencing of DNA samples was carried out commercially by MWG Eurofins Genomics (Ebersberg, Germany). Plasmid DNA was prepared as recommended by MWG. For high copy number expression vectors, the MWG T7 sequencing primer was routinely used and for the pMTL23-type substrate vectors, the MWG M13-rev (-49) sequencing primer was typically used.

## 2.21 *In vivo* (MacConkey) recombination assay

The MacConkey assay (Blake 1993) was used to screen for *in vivo* recombination activity. *E.coli* cells (DS941 strain, *galK*<sup>-</sup>) containing a low level resolvase expression plasmid (section 2.7.3) were made competent (section 2.8), usually using the CaCl<sub>2</sub> method and transformed (section 2.8.1) with ~0.01 µg *in vivo* recombination substrate plasmid, with a kanamycin resistance gene and two recombination sites in direct repeat flanking a *galK* gene.

Following transformation, cells were plated out on MacConkey indicator agar plates (Difco), with 2% galactose added and incubated overnight at 37°C. Plates were photographed (Canon EOS D30). A sample of cells was then scraped off the plates and grown overnight in L-broth. Plasmid DNA isolated from pelleted cells was then analysed by agarose gel electrophoresis following EtBr staining of the DNA.

Recombination of the substrate plasmid by resolvase (resolution) produces two circular DNA molecules, one containing the origin of replication and kanamycin resistance gene and the other the *galK* gene. The latter, having no origin of replication, is not maintained in the cell and is lost in subsequent cell divisions. MacConkey agar has no carbon source (other than the added galactose) and contains a pH indicator (2-methyl-3-amino-6-dimethylaminophenazine) which is yellow at pH >8.0 and red at pH <6.8. Successful resolution of the substrate plasmid forces the *E.coli* to metabolise amino acids in the agar as they are unable to metabolise galactose, resulting in an increase in pH and production of white colonies. Unresolved substrate with the *galK* gene, allows the *E.coli* to metabolise galactose, lowering the pH and producing red colonies (figure 2.3).



**Figure 2.3:** MacConkey assay.

(A) *E.coli* cells containing a pMS183 $\Delta$ -based 2-site recombination substrate were made competent and transformed with a pMS140-based low level resolvase expression plasmid. The transformants were then spread onto indicator plates containing galactose as a food source and an indicator dye which turns red in acidic conditions. The substrate has a kanamycin resistance gene and two recombination sites in direct repeat flanking a *galK* gene. In the presence of an active recombinase, recombination occurs and the *galK* gene is excised. The *galK*-containing resolved plasmid has no origin of replication and is lost. In the absence of *galK*, cells are unable to metabolise galactose, giving white colonies. (B) Red colonies formed in the presence of *galK* gene (unresolved substrate). (C) White colonies formed in the absence of *galK* gene (resolved substrate).

## 2.22 Purification of Proteins

Resolvase and TALER proteins were purified using a series of wash steps followed by affinity column purification (GE Healthcare, His Gravitrap affinity columns<sup>TM</sup>) and dialysis. Expression plasmids with T7 inducible promoters (section 3.2, table 2.5) were induced with 200 mM isopropylthio- $\beta$ -galactoside (IPTG) in the *E.coli* strain BL21(DE3)[pLysS]. Cells from 400 ml culture were harvested and broken by sonication (Sonics & Materials, VibraCell<sup>TM</sup>) by three sonication pulses of 15 second duration of cells, while kept cool on ice, in fractionation buffer 1 (25 mM inorganic sodium phosphate (NaPi; pH7.6) (a 3:1 ratio of di-sodium phosphate ( $\text{Na}_2\text{HPO}_4$ ) to monosodium phosphate ( $\text{NaH}_2\text{PO}_4$ )), 200 mM NaCl, 1 mM EDTA, 100  $\mu\text{g/ml}$  PMSF). Centrifugation at 9000 rpm for 5 mins at 4°C was then carried out and the resulting pellet re-suspended in fractionation buffer 1. Centrifugation was repeated and the sample was re-suspended in fractionation buffer 2 (25 mM NaPi, 1 M NaCl, 1 mM EDTA, 100  $\mu\text{g/ml}$  PMSF). This was then repeated for a final time in fractionation buffer 3 (25 mM NaPi, 2 M NaCl, 1 mM EDTA, 100  $\mu\text{g/ml}$  PMSF).

The fractionated samples were then added to nickel affinity (His Gravitrap) columns and washed sequentially with 10 ml buffer Q (25 mM NaPi, 20 mM imidazole, 1 M NaCl, 6 M urea), 10 ml Buffer B (25 mM NaPi, 20 mM imidazole, 2 M NaCl), Buffer W (25 mM NaPi, 100 mM imidazole, 2 M NaCl), 5 ml Buffer C (25 mM NaPi, 200 mM imidazole, 2 M NaCl), 5 ml Buffer D (25 mM NaPi, 400 mM imidazole, 2 M NaCl), 5 ml Buffer E (25 mM NaPi, 1.6 M imidazole, 800 mM NaCl), and 5 ml Buffer G (25 mM NaPi, 400 mM imidazole, 6 M urea). Purification of Tn3-TALERS was also carried out by nickel column wash steps; however, buffers Q and G were not used.

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) was then used to determine which buffer the protein eluted in (section 2.24). Protein samples were then further purified and concentrated by dialysis using dialysis tubing. Samples were immersed 1 L each of two buffer changes; first

for 16 h in buffer 1A (2 M NaCl, 10 mM NaPi, 0.1 mM EDTA, 200 mg/ml PMSF, 1 mM DTT), followed by transfer to the final storage buffer, buffer 2 (1 M NaCl, 10 mM Tris/HCl (pH 7.5), 0.1 mM EDTA, 50% w/v glycerol).

## **2.23 Denaturing polyacrylamide gel electrophoresis**

SDS-PAGE (Laemmli, 1970) was used to analyse protein samples. The protein samples were run through two gel layers of different ionic strength and acrylamide percentage. The first gel (stacking gel) was typically 5% acrylamide (5% acrylamide/bisacrylimide, 125 mM Tris/HCl (pH 6.8), 0.1% SDS, 0.1% APS, 0.2% TEMED), the second gel (resolving gel) was typically 15% (15% acrylamide/bisacrylimide, 375 mM Tris/HCl (pH 8.8), 0.1% SDS, 0.1% APS, 0.2% TEMED). The resolving gel was poured between clamped, sealed glass plates and covered with isopropanol to exclude oxygen to aid polymerisation. When set, the isopropanol was washed off, the stacking gel solution was poured in and a well-forming comb was inserted. After 30-40 minutes of polymerisation, the comb was removed and the gel was inserted into a vertical electrophoresis kit. The gels were run in Laemmli electrophoresis buffer (25 mM Tris base, 250 mM glycine, 0.1% SDS).

Prior to loading of protein samples, Laemmli loading buffer (50% (v/v) glycerol, 5% SDS, 200 mM Tris/HCl (pH 6.8), 0.1 mM DTT) was added at a ratio of 1:5 and the samples were heated to 100°C for 5 minutes to denature the proteins. Samples were then loaded onto the gel and run at 30-40 mA (200 V) for 2.5 - 3 hours.

## 2.24 *In vitro* cleavage assays

*In vitro* cleavage assays were carried out in a reaction buffer with a final concentration of 50mM tris/HCl (pH8.2), 0.1mM EDTA, 40% Ethylene Glycol, 100mM NaCl and 25µg plasmid DNA. 2.2µl protein was added to the reaction mix and incubated at 37 °C for 2 hr. Reactions were stopped by the addition of 5 µl SDS loading buffer with 1 µl Protease K added prior to running on an agarose gel (section 2.13.i).

## 2.25 Radioactive labelling of DNA for binding assays

DNA was labelled using either an  $\alpha$ -<sup>32</sup>P [dCTP] or a  $\gamma$ -<sup>32</sup>P [ATP] source,  $\alpha$ -<sup>32</sup>P [dCTP] was used for labelling of double stranded DNA fragments from plasmid DNA and  $\gamma$ -<sup>32</sup>P [ATP] was used for labelling of top-strand synthetic oligonucleotides prior to annealing. For labelling with  $\alpha$ -<sup>32</sup>P [dCTP], Klenow DNA polymerase was used. Labelling reactions were used for double stranded annealed oligonucleotides or digested plasmid DNA, and contained 1x NEB buffer 4 (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium Acetate, 1 mM DTT), 10-20 µM of the 3 unlabelled dNTPs (dATP, dGTP and dTTP), 0.37 MBq of  $\alpha$ -<sup>32</sup>P [dCTP] and 1-2 units of Klenow polymerase. The labelling reaction was carried out for 10 mins at room temperature. The 10 min incubation was followed by addition of unlabelled (dATP, dGTP, dCTP and dTTP) to a final concentration of 40 µM and a further 10 min incubation step.

For  $\gamma$ -<sup>32</sup>P labelling, single stranded oligonucleotides were purified by gel extraction followed by ethanol precipitation (section 2.12). The concentration of the oligonucleotides was then measured and the top strand was 5' end labelled with  $\gamma$ -<sup>32</sup>P using T4 polynucleotide kinase (T4-PNK). Labelling reactions were typically carried out in 20 µl and contained 0.37 MBq  $\gamma$ -<sup>32</sup>P [ATP], 1x polynucleotide kinase (PNK) buffer, 1 µM oligonucleotides and 10

Units of T4-PNK. The reaction was carried out for 1 h at 37°C, followed by heating to 95°C for 5 mins. 40 µl of TE buffer was then added to the 20 µl reaction and the solution of labelled oligonucleotide was filtered through a column to remove unbound label (GE Healthcare, Illustra Micro Spin G-25 columns<sup>TM</sup>). An excess of bottom strand oligonucleotide (typically 2.2 x), 20 µl TE buffer and 100 mM NaCl was then added and the sample was heated to 95°C for 10 mins, followed by slow cooling to room temperature to anneal. Ethanol precipitation was then carried out (section 2.12).

## **2.26 Gel purification of labelled DNA**

For purification of labelled DNA, 1x Ficoll 'stop mix' (20% Ficoll, 0.1% SDS, 0.01% bromophenol blue) was added in a 1:5 ratio and the samples were run on a preparative 6% Tris-borate polyacrylamide gel (section 2.13.ii). The gel was then routinely stained with 0.5 µg/ml EtBr (section 2.16).

## **2.27 Crush and soak**

DNA from acrylamide gel chips cut out from a gel was extracted by crushing the gel chip in 500 µl TE Buffer. Following overnight incubation with shaking at 37°C, 1000 rpm (Eppendorf Thermomixer Compact), the sample was filtered to remove any remaining gel by centrifugation at 10,000 rpm for 2 min using 0.2 µM Costar filter columns. The flow-through was collected for ethanol precipitation (section 2.11) or phenol/chloroform extraction (section 2.27).

## **2.28 Phenol chloroform extraction**

For extraction of DNA from acrylamide gels subsequent to the crush and soak protocol, 100 µl of a 1:1 ratio of phenol/chloroform was added. Following



centrifugation (13,000 rpm for 1 min) the aqueous phase was transferred to a fresh Eppendorf tube, a further 100  $\mu$ l of chloroform was added and mixed by vortexing, the centrifugation step was then repeated. The aqueous phase was again transferred to a new Eppendorf tube prior to ethanol precipitation of the DNA.

## **2.29 *In vitro* binding assays**

Binding was typically carried out in a 20 $\mu$ l reaction mix, containing 1x ficoll loading buffer (50% v/v ficoll, 200 mM tris/HCl), 20% glycerol, 1 unit of Poly(deoxyinosinic-deoxycytidylic) acid and 200 nM labelled Site I DNA. 1 $\mu$ l of purified resolvase ( $\sim$ 30  $\mu$ M) was then added and the reaction mix was incubated at room temperature for 20 min, transferred to ice for 10 min and loaded onto a 5% polyacrylamide gel with glycerol (section 2.13.ii). Gels were typically run at 240V (40mA) for 3 hr prior to the gel being dried and exposed overnight to image plates. Plates were scanned using typhoon imager.

## **2.30 Crystal structure models**

For crystal structure models where more than one structure is used, the structures were superimposed using Pymol molecular graphics program. Models were typically aligned to the E-helices of resolvase.

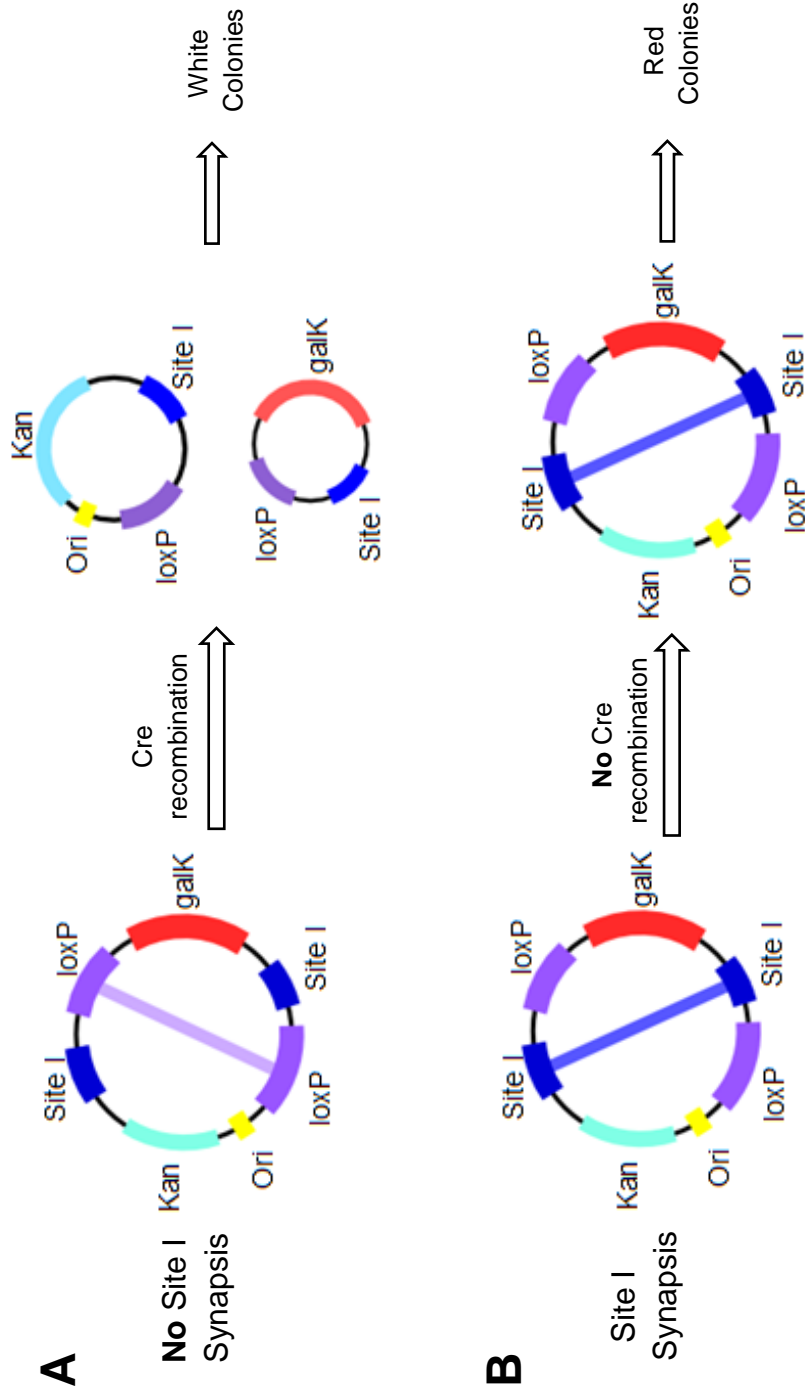
## Chapter 3: Roles of the E-helix in the DNA-binding and synapsis of Sin resolvase

### 3.1 Introduction

WT Sin resolvase binds as a dimer to site I of its recombination site *resH* (section 1.4), but two dimer sites do not interact to form a synapse except in the presence of the accessory sites, the accessory bending HU-like protein and negative supercoiling (section 1.4). In contrast, the activated Sin mutant Q115R binds to isolated site I and forms a synaptic tetramer complex.

In order to find other mutants of WT Sin that are synapsis-competent, an assay was carried out by Sally Rowland which was designed to screen for mutants of resolvase which showed higher levels of synapsis than the WT protein (Sally Rowland, unpublished results). This assay, similar to that described in Mouw et al. 2008, was carried out using a library of recombination-deficient Sin mutants, with the catalytic serine residue mutated to an alanine (S9A), which were screened by transforming into cells expressing Cre recombinase and containing a substrate plasmid with two *res* site I sites, and two Cre *loxP* sites, flanking a *galK* gene (figure 3.1). In the absence of synapsis by Sin, Cre recombinase recombines the *loxP* sites, the *galK* gene is lost and white colonies are produced on MacConkey indicator plates. In the presence of synapsis-competent Sin mutants, synapsis of the Sin *res* sites blocks Cre recombination, giving red colonies on MacConkey plates.

The activated mutant Q115R was identified multiple times in this screen, both on its own and in conjunction with additional mutations. Interestingly, in addition to these mutants, truncated variants of Sin resolvase were also isolated from this assay. These variants had premature stop codons, and contained activating mutations and were deduced to be creating truncated proteins, with the catalytic N-terminal domain absent.



**Figure 3.1: Constructs used in ‘synapsis-up’ screen**

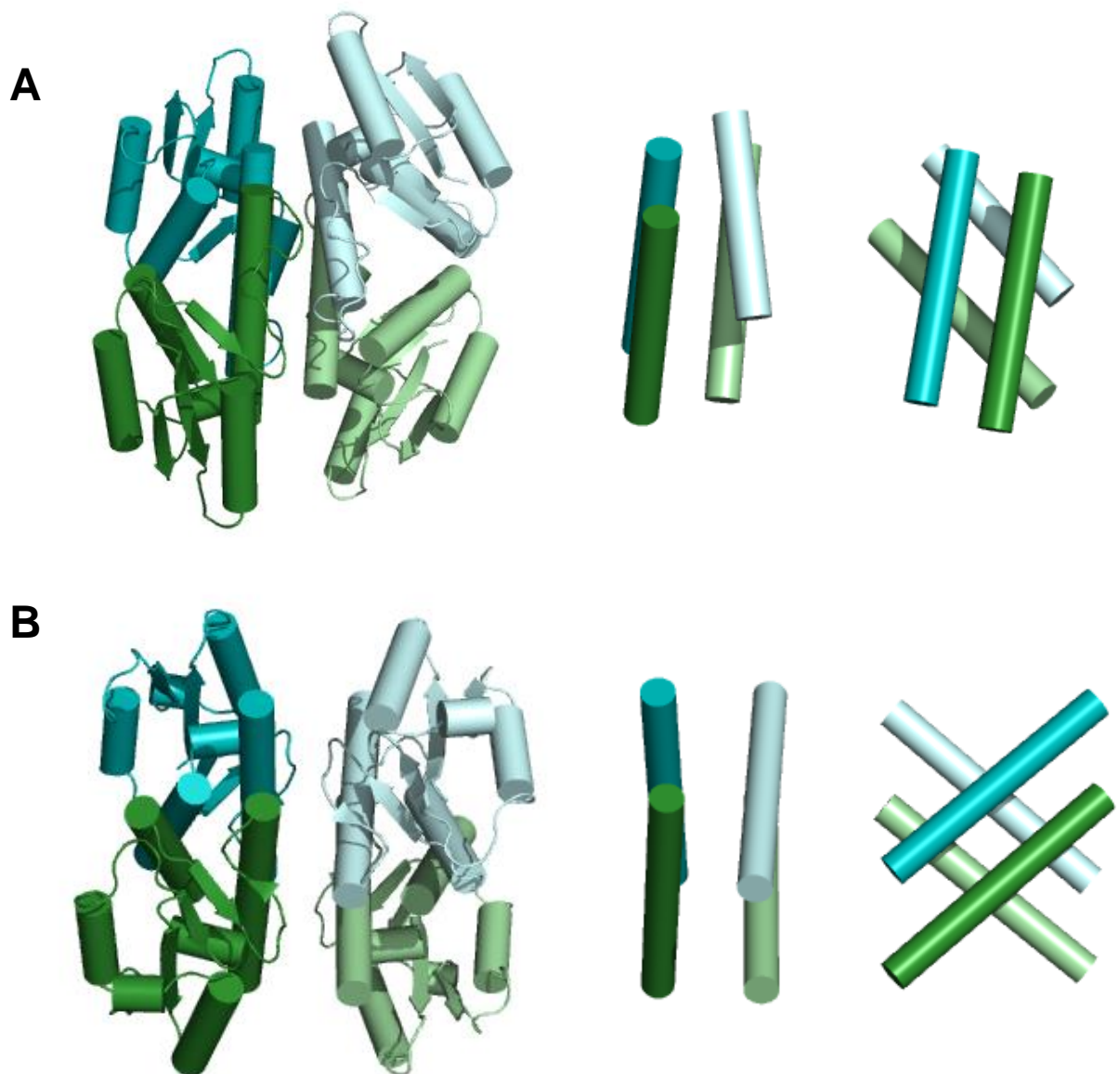
Plasmids utilised in the screen had *Sin* *res* sites and Cre *loxP* sites, flanking a *galK* gene, an origin of replication and a kanamycin resistance gene for selection. **(A)** In the absence of synapsis by *Sin*, Cre recombines its *loxP* sites, leading to loss of the *galK* gene and production of white colonies on indicator plates. **(B)** In the presence of synapsis-competent, but recombination deficient (*S9A*) *Sin*, Cre access to its *loxP* sites is blocked, the *galK* gene is maintained, and red colonies are produced on indicator plates.

These truncated protein variants, (termed 'tadpoles'), due to the positive signal obtained in the assay previously described (Sally Rowland unpublished results) were hypothesised to be forming synaptic complexes, despite the absence of the N-terminal domain (~100 residues). It was hypothesised that these 'tadpole' constructs bind to and dimerise on site I DNA in the same way as WT Sin and that the E-helices make a synapsis interface like that seen in the core of the crystal structures (figure 3.2A). This same 4-helix bundle is present in other members of the recombinases family, including  $\gamma\delta$  resolvase (figure 3.2B). The conserved nature of this helix-bundle suggests that it may form a 'synapsis module', the minimum architecture required to hold the dimers together and permitting subunit rotation to take place following synapsis.

As the 'tadpole' proteins are WT in sequence, but are nonetheless capable of synapsis unlike the full-length WT protein, it can also be deduced that the catalytic N-terminal domains of Sin (and  $\gamma\delta$ ) resolvase have a negative regulatory role, blocking E-helix interactions with each other (and thus synapsis) unless the correct synapsis complex is formed in the full WT system.

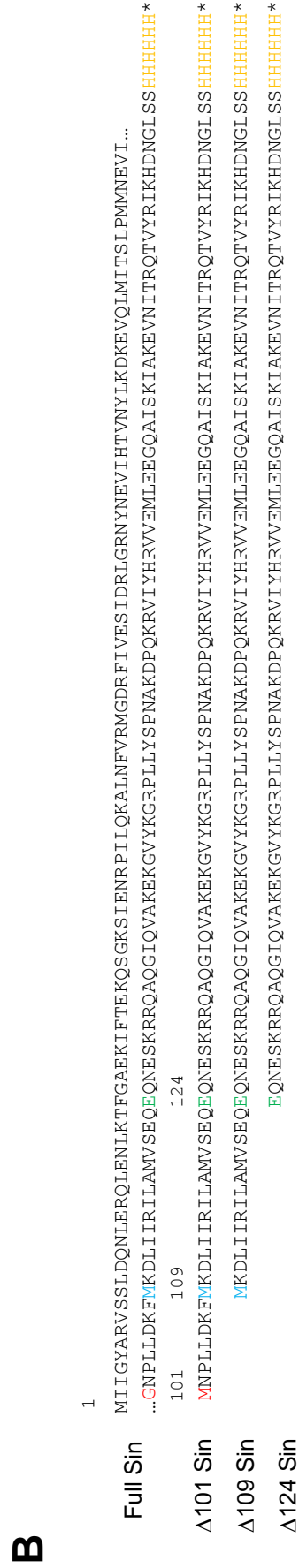
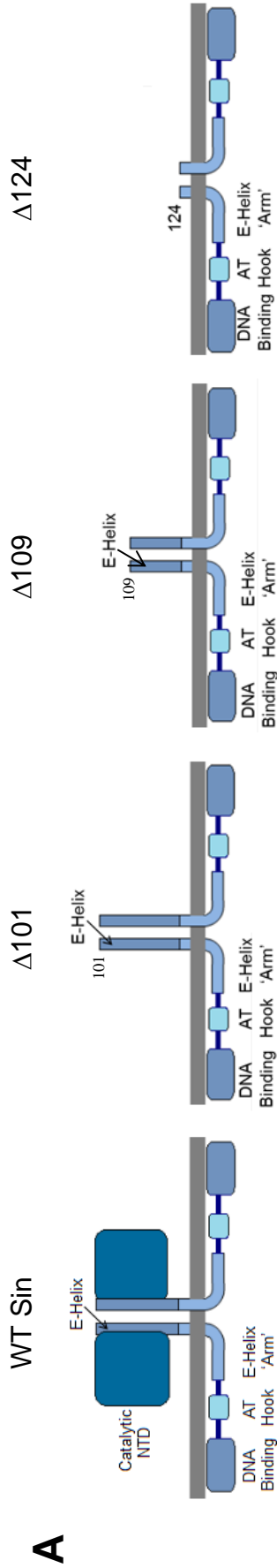
Following on from the discovery of these 'tadpoles', this chapter will discuss the purification and testing of three lengths of truncated Sin variants, with deletions of the first 100 residues ( $\Delta 101$ ), 108 residues ( $\Delta 109$ ) and 123 residues ( $\Delta 124$ ). The  $\Delta 109$  protein was isolated in the initial synapsis-up screen, the longer,  $\Delta 101$  and the shorter  $\Delta 124$  protein lacking the E-helix region were created to determine the importance of the E-helix to synapsis (figure 3.3).

The properties of these 'tadpole' proteins were determined using *in vitro* binding and synapsis assays, as described by Olorunniji et al (2008) and Rowland et al (2009).



**Figure 3.2: Resolvase ‘synapsis module’**

Crystal structures illustrating the catalytic domains and E-helices of (A) Sin resolvase (3PKZ, Keenholtz et al. 2011) (B)  $\gamma\delta$  resolvase (1ZR4, Li et al. 2005) and the 4-helix bundle at the core of the catalytic tetramers forming the hypothesised ‘synapsis module’.



**Figure 3.3: Truncated Sin enzymes:**

(A) Cartoons showing: Full-length Sin dimer, ΔM101 Sin dimer with N-terminal truncation of residues 1-100, ΔM109 Sin dimer with N-terminal truncation of residues 1-108 and ΔM124 Sin dimer with N-terminal truncation of residues 1-123.

(B) Sequence of full-length Sin and the truncated 'tadpoles' Δ101, Δ109 and Δ124, with positions 101 (red), 109 (blue), 124 (green) and the N-terminal Histidine tag (orange) highlighted.

As the ‘tadpoles’ are predicted to recognise the site I DNA in the same way as the ‘activated’ full-length Sin/Tn3/ $\gamma\delta$  proteins (see section 1.2), the interaction between the DNA within the minor groove of the site and the ‘arm’ region of the protein should be an important one. As such, the effect of ‘swapping’ the arm region of full-length and ‘tadpole’ Sin proteins with the equivalent residues from Tn3 resolvase which naturally interact with a very different central sequence was tested. It was predicted that these hybrid proteins would fail to interact with the Sin site I, but may have the ability to form a synapse at a hybrid site I, with the relevant residues replaced with sequence from Tn3 site I.

The properties of the WT ‘tadpole’ variants were also compared with point mutants within the E-helix, selected based on the available crystal structure data, to be informative about the mechanism of synapsis. These assays aimed to determine whether the synapses formed by the tadpole variants are structurally similar to the synapses formed by the assembly of the four-helix bundle seen in  $\gamma\delta$  and Sin crystal structures. These mutations were also introduced into the full-length Q115R Sin protein, with the hypothesis that they will have an effect on binding and recombination consistent with the effects on ‘tadpole’ synapsis. Q115R was utilised for this purpose, as it is the single strongest activating mutation in Sin, and is known to be synapsis-competent *in vitro*.

### **3.2 Sin ‘tadpole’ design and construction**

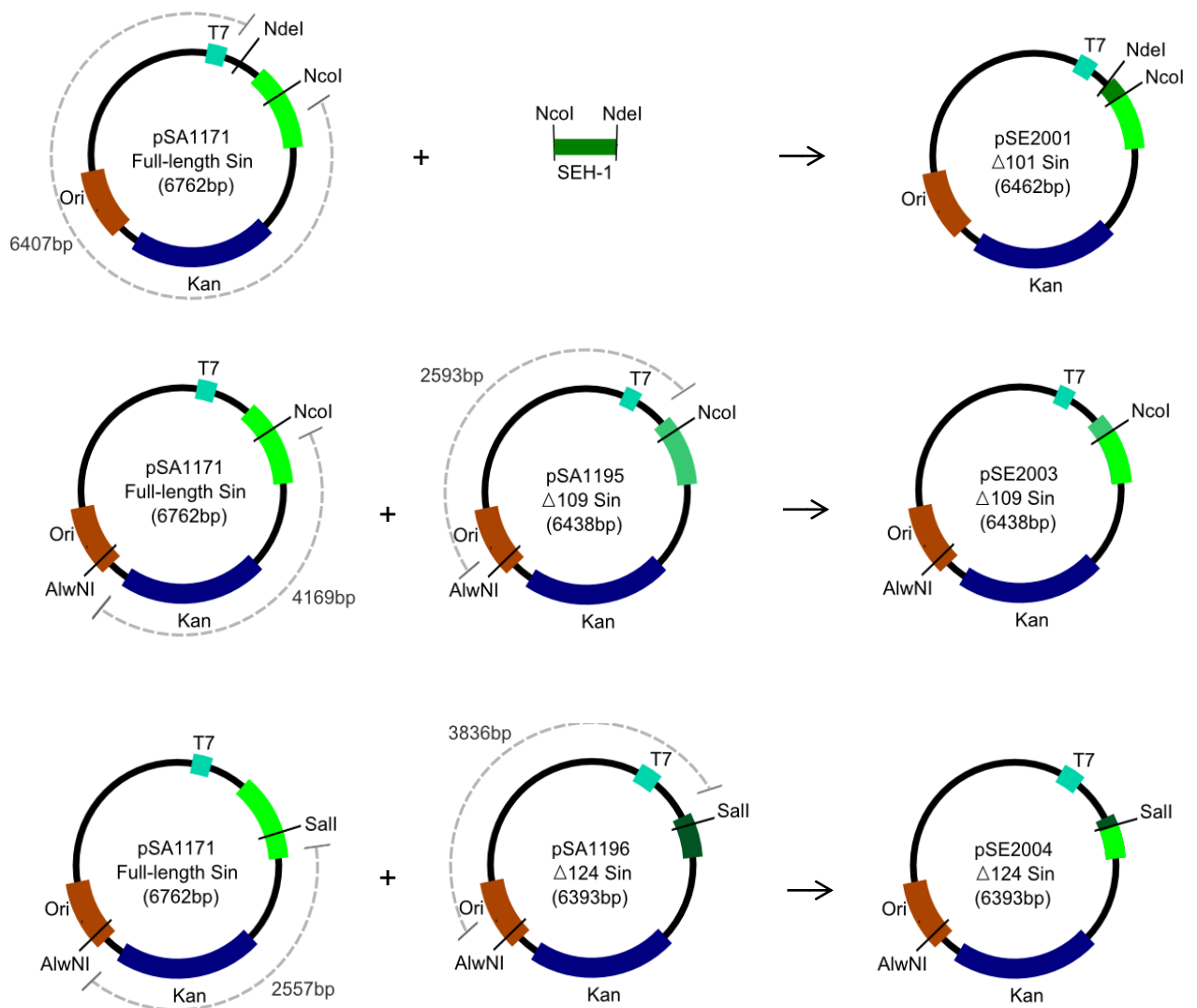
The three ‘tadpole’ proteins utilised in this study,  $\Delta 101$ ,  $\Delta 109$  and  $\Delta 124$  were created by fragment-swap or by insertion of oligonucleotides into pre-existing vector plasmids. High copy number plasmids (pSA1101-type with inducible T7 promoters, see table 2.5) were constructed for *in vitro* expression of Sin resolvase and the ‘tadpole’ proteins. The WT  $\Delta 101$  protein was created by ligation of oligonucleotide SEH1 (see table 2.4) into T7 inducible high copy

number expression vector pSA1171 (full-length Sin, see table 2.5) between the restriction sites NdeI and NcoI. The WT  $\Delta 109$  expression vector was created by AlwNI–NcoI fragment swap between the expression vectors pSA1171 and pSA1195 ( $\Delta 109$  V138A Sin, see table 2.5) and the shortest truncated variant,  $\Delta 124$  was created by fragment swap between pSA1171 and pSA1196 ( $\Delta 124$  V138A Sin, see table 2.5) as illustrated in figure 3.4. Each of these proteins has a histidine tag (his<sub>6</sub>) C-terminal to the DNA-binding domain as shown in figure 3.3. Following creation of these vectors as described, the plasmids were transformed into the *E.coli* strain BL21 (DE3) [pLysS] (see table 2.5) using the chemically competent transformation method described in section 2.8-2.9. These cells were then grown to mid-log phase, as determined by spectrophotometry, and induced by addition of IPTG (as in section 2.22).

Protein induction was visualised by SDS-PAGE, with un-induced and induced samples visualised by staining with coomassie brilliant blue staining (figure 3.5A). The cells were pelleted and the samples fractionated (as described in section 2.22). The fractions were visualised by SDS-PAGE and the best sample was selected for purification by nickel column chromatography using the C-terminal his<sub>6</sub> tag (section 2.22).

Following nickel column purification, further SDS-PAGE was used to determine the best sample to further purify and concentrate. This sample was then dialysed (as described in section 2.22) to concentrate the sample and yield a sample in 50% glycerol buffer for storage at -70°C (figure 3.5B). The concentrations of the protein samples were determined by visualisation on polyacrylamide gels, followed by normalisation to the protein of lowest concentration.

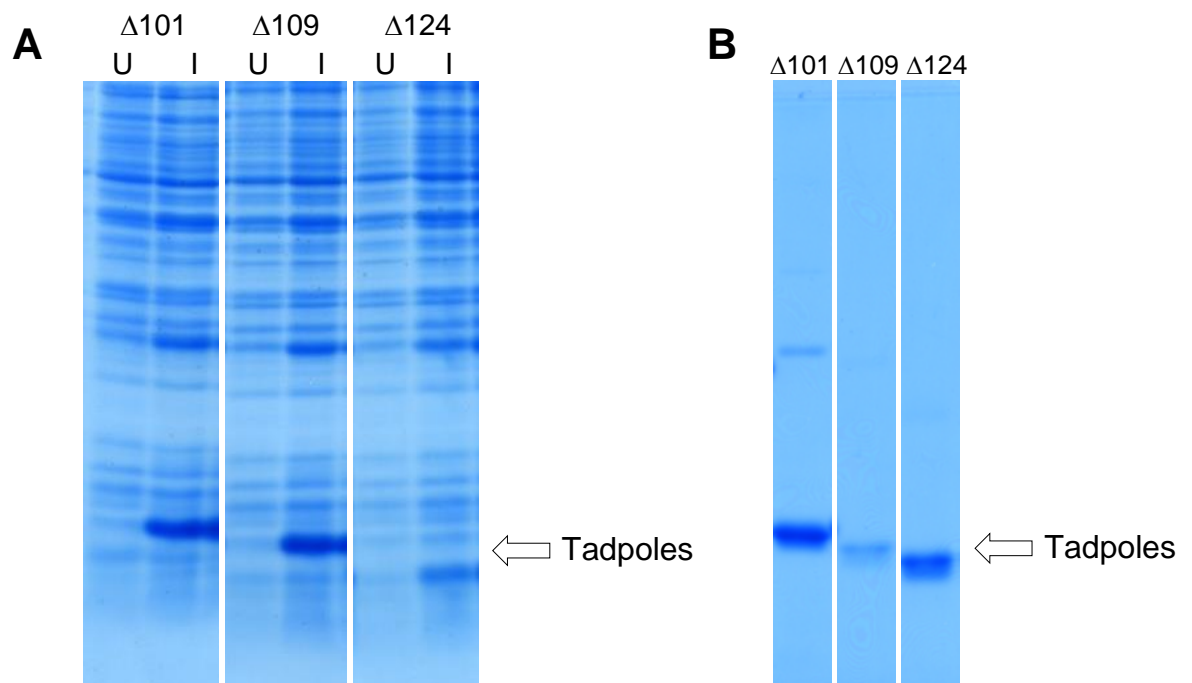




**Figure 3.4: Expression plasmid construction**

Construction of high copy number IPTG-inducible ‘tadpole’ expression vectors.

**(A)**  $\Delta 101$  construction by ligation of the oligonucleotide SEH1 into pSA1171 (NdeI-NcoI). **(B)**  $\Delta 109$  construction by fragment-swap between pSA1171 and pSA1195. **(C)**  $\Delta 124$  construction by fragment-swap between pSA1171 and pSA1196.



**Figure 3.5: SDS-PAGE**

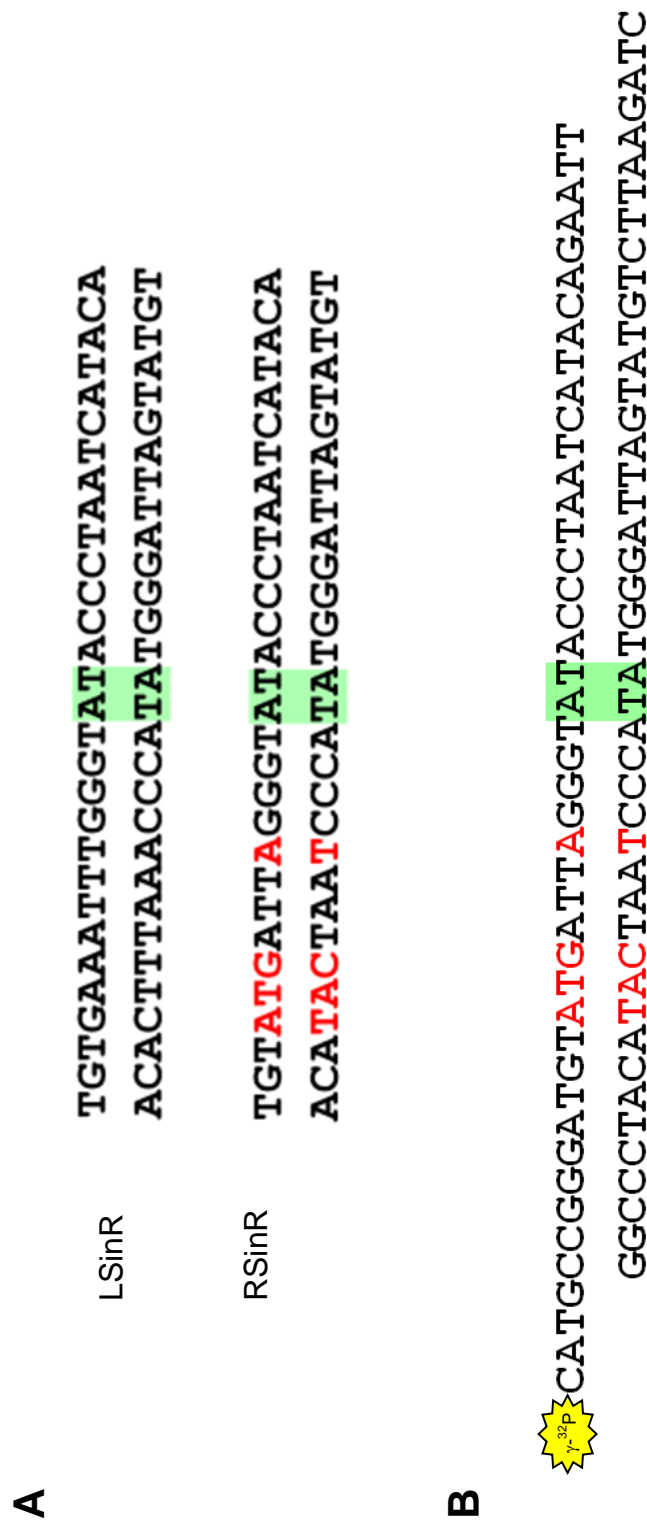
(A) Protein samples prior to (U) and following induction (I) with IPTG, showing expression of the 'tadpole' proteins. (B) The 'tadpole' constructs following column purification and dialysis.

### 3.3: Binding and synopsis of *res* site I

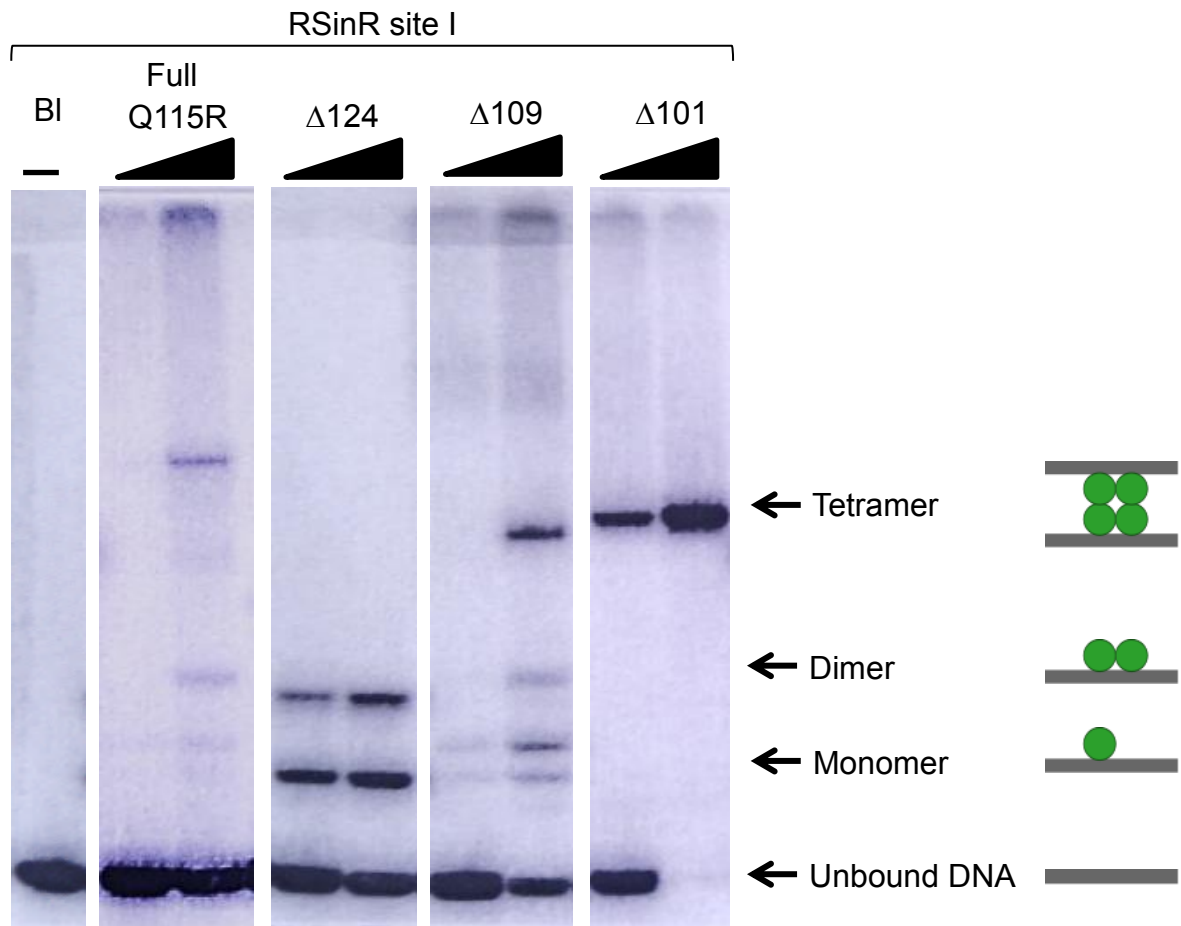
The 'tadpole' proteins were tested for binding and synopsis to *res* site I. For the purpose of these experiments, a symmetrised Sin site I variant (RSinR) was utilised, with the left hand side of the site symmetrised to the right side of the site, as previous research has indicated that the binding affinity for the right half of the site is better than that of the left half (Sally Rowland and Martin Boocock unpublished data) (figure 3.6A). This RSinR substrate was thus predicted to be a better substrate for these *in vitro* assays, as it should bind to the site with greater affinity than the LSinR site. Linear substrates, 45 bp in length, with the 30 bp RSinR substrate in the centre, were ordered as oligonucleotides (SEH 52-T and SEH 52-B from table 2.4) from Eurofins MWG Operon (figure 3.6B).

For visualisation of protein-DNA complexes *in vitro*, purified top strand oligonucleotide (SEH 52-T) was 5' end-labelled with  $\gamma$ -<sup>32</sup>P ATP (section 2.24) and annealed to the bottom strand (section 2.18). Binding assays were typically carried out using gels with 7% acrylamide/bis-acrylamide (30% w/v, 37.5:1) and 10% glycerol, run with TBE buffer (section 2.13.iii). Glycerol was added to the gels as it was determined that this increased the stability of the protein-DNA complexes on the gel. Gels were routinely run at 4 °C for 4-5 hrs at 200 V (~15 W). These binding assays typically utilised unbound DNA as a control for band-shift of complexes.

These assays used full-length Q115R Sin as a control, as this activated mutant has been demonstrated previously to form monomer, dimer and tetramer structures on polyacrylamide gels. Three sizes of retarded protein-DNA complexes were visualised in these assays (figure 3.7). The full-length Q115R Sin produced the predicted three bands, as previously established. The longest of the truncated proteins, the  $\Delta$ 101 'tadpole' formed only the slowest migrating band, hypothesised to be a synaptic tetramer structure, based on the band-shift position relative to the full-length protein complexes.



**Figure 3.6: Sin site I res sequence**  
**(A)** AT-centred res site I (LSinR) (top) and site I symmetrised to the right hand side (RSinR) (bottom) **(B)** RSinR site oligonucleotides SEH52-T and SEH52-B, 5' end labelled with  $\gamma$ - $^{32}\text{P}$  ATP for *in vitro* binding assays.



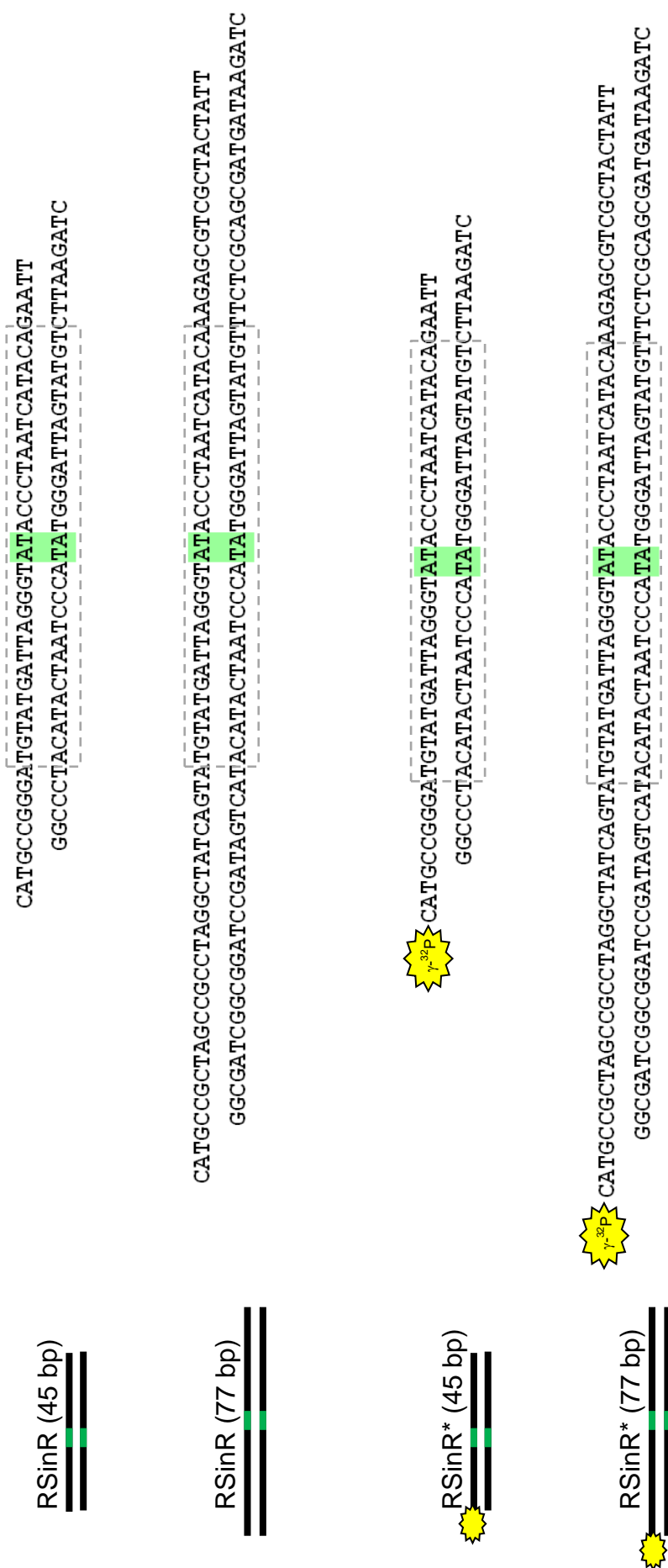
**Figure 3.7: *In vitro* binding assay**

*In vitro* binding assay with  $\gamma$ -<sup>32</sup>P labelled RSinR site I DNA substrates bound by full-length Q115R Sin and by the truncated ‘tadpole’ constructs;  $\Delta$ M101,  $\Delta$ M109,  $\Delta$ M124 and a no protein control (BI).

The  $\Delta 109$  'tadpole' formed three complexes on the gel, hypothesised to be monomer, dimer and tetramer complexes. An additional, faster migrating band is also present on the gel, which may be an altered conformation of the tadpole. The shortest  $\Delta 124$  'tadpole' formed the fast and intermediate running complexes but not the slowest, indicating that it is able to form the hypothesised monomer and dimer complexes, but not the tetramer structure hypothesised as the slowest migrating complex on the gel.

The slowest migrating of these complexes is thus hypothesised to contain two DNA molecules bound by a tetramer of Sin, the intermediate complex is hypothesised to have one molecule of DNA bound by a dimer of Sin and the fastest running complex is hypothesised to contain one molecule of DNA bound by a monomer of Sin.

If the hypothesis that the slowest migrating bands seen for the  $\Delta 101$  and  $\Delta 109$  'tadpoles' in figure 3.10 represents a synapse analogous to that seen in the full-length Q115R Sin is correct, then the complex must contain two molecules of DNA bound by four subunits of resolvase. In order to test the validity of this hypothesis, a further binding assay was designed, utilising two different lengths of DNA (45 bp and 77 bp) with the 30 bp RSinR *res* site I at the centre (figure 3.8).



**Figure 3.8: RSinR binding substrates**

The substrate utilised in the 'mixed length' binding assay. Unlabelled DNA of 45 bp (RSinR 45) and 77 bp (RSinR 77) and the same sequences, labelled at the 5' end of the top strand with  $\gamma$ -<sup>32</sup>P (RSinR\* 45) and (RSinR\* 77). The sites are represented by dotted boxes.

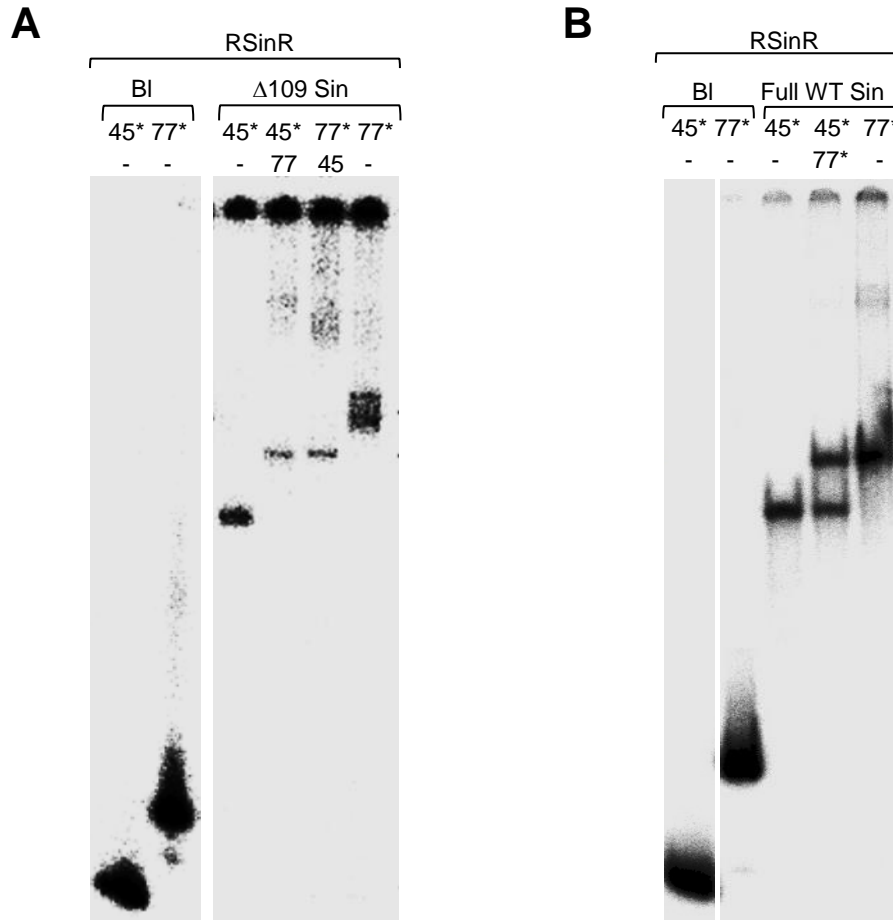
Each DNA molecule (45 bp or 77 bp) was either labelled with  $\gamma^{32}\text{P}$  (\*) or left unlabelled. Linear DNA molecules of 45 bp and 77 bp (45 + 77 and 45\* + 77\* respectively) and the  $\Delta 109$  'tadpole' were utilised for this experiment. The full-length WT Sin, similar to WT Tn3 resolvase, has been shown to form only dimer *in vitro* (Rowland et al 2009) and would be expected to form two distinct bands in this assay, one with 45\* and one with 77\*.

This assay tested the  $\Delta 109$  protein, with full-length WT Sin as a control (figure 3.9). The full-length WT, as expected, forms dimer complex on the gel, visualised as two bands (a 45\* DNA molecule bound by a dimer of Sin or a 77\* DNA molecule bound by a dimer of Sin). The  $\Delta 109$  'tadpole' forms the three bands anticipated for the presence of a tetramer like that seen in full-length Q115R Sin; two 45\* DNA molecules bound by a tetramer of Sin, a 45\* and a 77 DNA molecule, or a 45 and a 77\* DNA molecule bound by a tetramer of Sin or two 77\* DNA molecules bound by a tetramer of Sin. This result confirmed that the large complex formed by the  $\Delta 109$  'tadpole' proteins contains two molecules of DNA as hypothesised for the predicted tetramer complex.

### **3.4 Sequence selectivity of *res* site I**

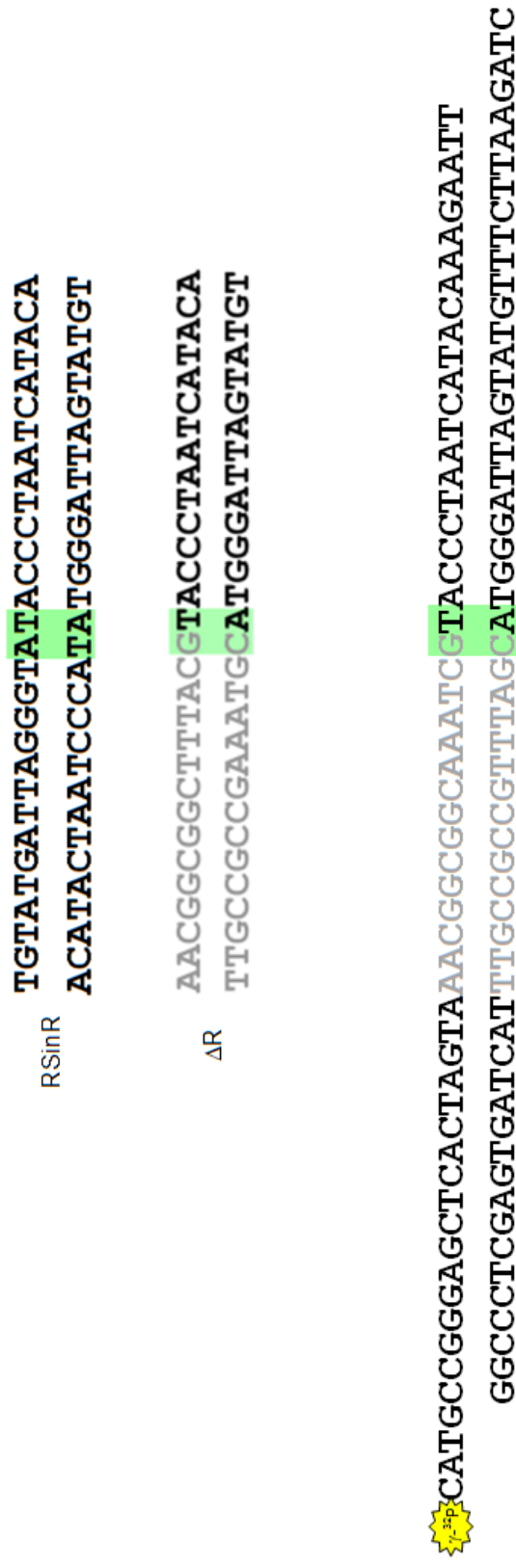
Having demonstrated that the 'tadpole' constructs are forming a synapse analogous to that formed by the full-length activated mutant Q115R Sin, it was next desirable to determine whether there is any sequence selectivity at the centre of the site, and what effect if any changing this sequence would have on binding and synapsis of this site. In order to test the importance of the site sequence to binding and synapsis, a half-site ( $\Delta R$ ) was designed (figure 3.10), with the right hand side of Sin as normal, and the left hand side of the site altered to residues selected to be as different from the right hand-





**Figure 3.9: *In vitro* band-shift assay**

Assay showing binding of  $\Delta 109$  Sin 'tadpole' to RSinR site I substrates of 45 bp and 77 bp.  $\gamma^{32}\text{P}$  labelled site I (45\* & 77\*) and non-labelled site I (45 & 77) were used. **(B)** Band-shift assay showing binding of full-length Sin to RSinR site I substrates. No protein controls (BI) are also shown.



**Figure 3.10: ΔR site I res sequence**

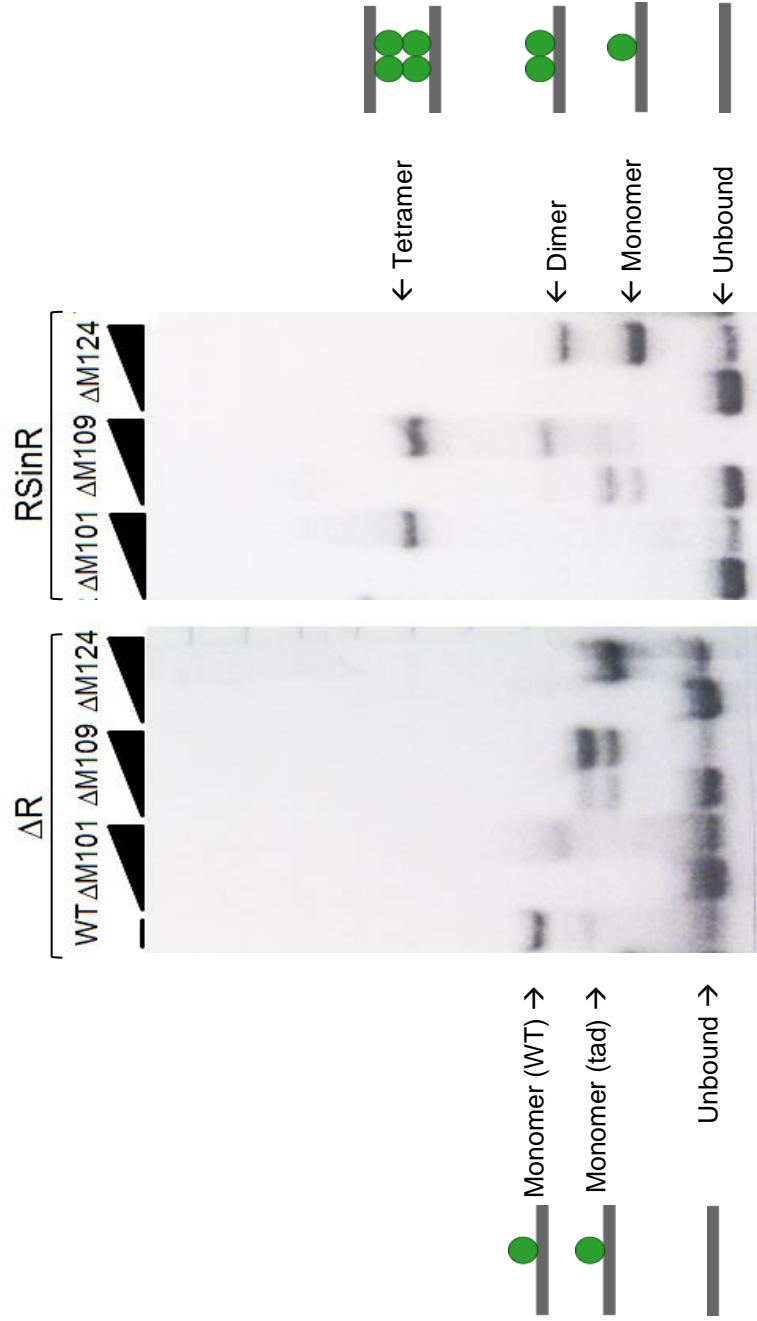
**(A)** AT centred symmetrised res site I (RSinR) (top) and half-site I (ΔR) with nucleotides designed to discourage binding (grey) to the left half of the site (bottom) **(B)** RSinR site oligonucleotides SEH52-T and SEH52-B, 5' end labelled with  $\gamma$ -<sup>32</sup>P ATP for *in vitro* binding assays.

side sequence as possible, a sequence predicted to be an undesirable Sin binding sequence. This sequence was ordered from MWG as top and bottom strand oligonucleotides (SEH 51-T and SEH 51-B, table 2.4). If, as hypothesised, the sequence of the site is important to binding, the  $\Delta R$  site would be predicted to be bound by a monomer of Sin at the right hand side of the site only.

The  $\Delta R$  top strand oligonucleotide was 5' end  $\gamma$ - $^{32}\text{P}$  ATP labelled and annealed to the bottom strand as described in section 2.24. Binding and synapsis assays were then carried out as with the RSinR assays previously discussed. Polyacrylamide gel electrophoresis under the conditions described in section 3.3 was carried out to test binding to the  $\Delta R$  site I (figure 3.11). It was observed that the  $\Delta R$  site as hypothesised is able to form only monomeric interactions with Sin, illustrating that both the right and left halves of the site are required for dimer and synapse formation, such as that seen with the RSinR substrate.

Once it had been determined that both halves of the site are necessary for any interactions involving more than one Sin subunit, the effect of replacing the sequence at the centre of the site was tested. This was achieved by altering the central 10 bp of the 30 bp site from the Sin *res* site I residues to Tn3 sequence, creating an RTn3R substrate (figure 3.12).

As for the RSinR and  $\Delta R$  substrates, the RTn3R sequence was ordered from MWG as top and bottom strand oligonucleotides (SEH 50-T and SEH 50-B, table 2.4) If the sequence at the centre of the site is important to binding, then the alterations made to the sequence in the RTn3R substrate would be predicted to disrupt binding of the 'tadpole' constructs.



**Figure 3.11: *In vitro* binding to  $\Delta R$  Site I**


Binding of full-length WT Sin and truncated 'tadpole' constructs to the  $\Delta R$  site I (left) and binding of the same 'tadpoles' to the RSinR site (right).

**A**

TGTATGATTAGGGTATACCCTAATCATACA  
ACATACTAATCCCATATGGGATTAGTATGT

TGTATGATTATATATAATAATCATACA  
ACATACTAATAATAATATATATAGTATGT

**B**

 CATGCCGGGAGCTCACTAGTATGTATGATTATATATAATAATCATACAAGAATT  
GGCCCTCGAGTGATCATACATACTAATAATAATATATATAGTATGTTCTTAAGATC

**Figure 3.12: RTn3R site I res sequence**

**(A)** AT centred symmetrised res site I (RSinR) (top) and RTn3R Sin, with central nucleotides changed to Tn3 sequence (red) **(B)** RTn3R site oligonucleotides SEH50-T and SEH50-B, 5' end-labelled with  $\gamma$ -<sup>32</sup>P ATP for *in vitro* binding assays.

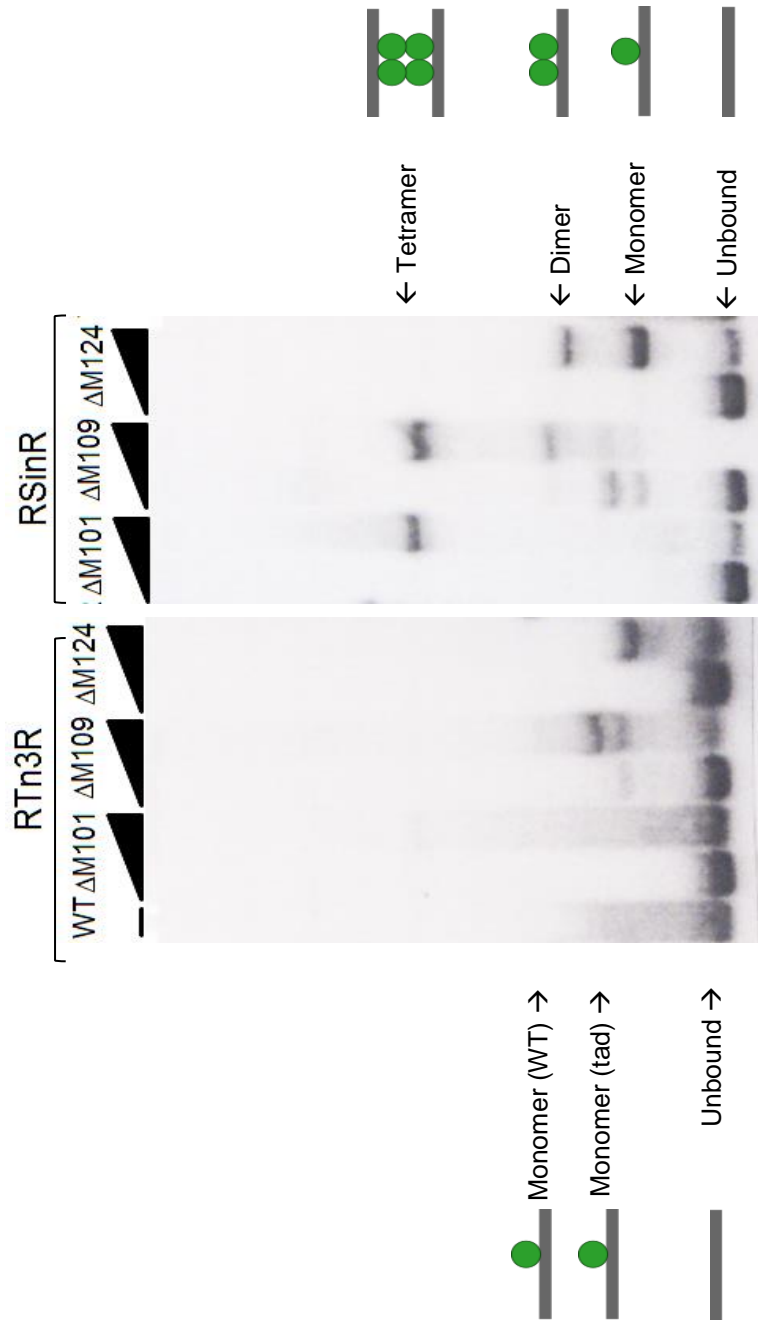
The RTn3R top strand oligonucleotide was 5' end  $\gamma$ -<sup>32</sup>P ATP labelled and annealed to the bottom strand as described in section 2.24. Binding and synapsis assays were then carried out as with the binding assays previously discussed for RSinR and  $\Delta$ R. Polyacrylamide gel electrophoresis under the conditions described in section 3.3 was carried out to test binding to the RTn3R site I (figure 3.13). This experiment used full-length WT Sin as a control for binding. It was observed that the RTn3R site behaved mostly as predicted, blocking the normal interactions seen with the full-length and 'tadpole' proteins on the RSinR site.

The full-length WT, the  $\Delta$ 109 and  $\Delta$ 124 'tadpoles' as hypothesised bound to the site as monomer only. The longest of the 'tadpole' constructs however shows evidence of a tetramer complex on the gel. This complex appears unstable however. The detrimental effect on the normal complexes seen on binding with the RTn3R substrate demonstrates that the central sequence, at the position where the 'arm' region of the E-helix crosses the DNA, is important to binding of the site and subsequent synapsis.

### **3.5: Sequence selectivity by the E-helix 'arm' of resolvase**

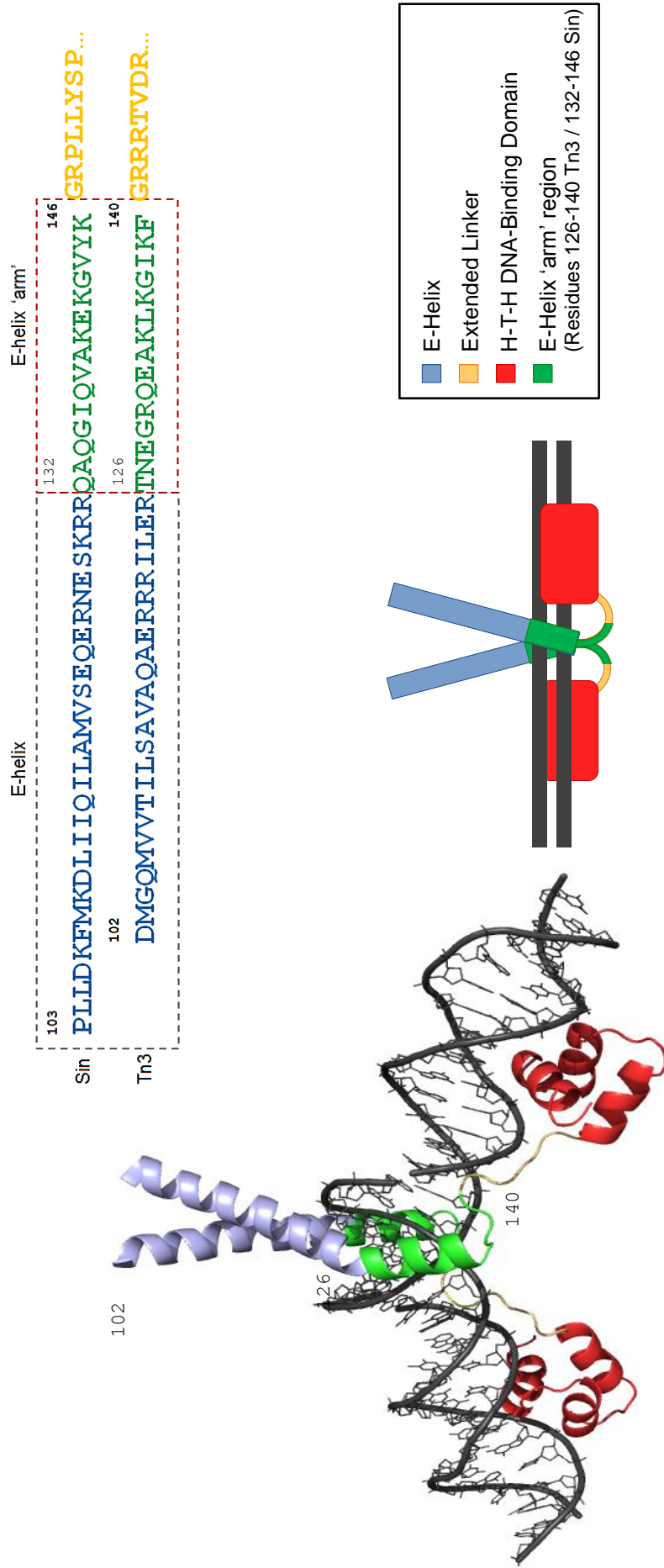
If this region, within the central 16 bp of the *res* site I sequence is making specific interactions with the 'arm' region of Sin resolvase, then hypothetically it should be possible to swap the 'arm' region of Sin resolvase with the equivalent residues from Tn3 resolvase (Sin residues 132-146, Tn3 residues 126-140) (figure 3.14). This should create a hybrid protein which is unable to stably bind to the RSinR substrate, but may have the ability to bind to and synapse the hybrid site.

Hybrid Sin proteins with the 'arm region of Tn3 were constructed by insertion of annealed oligonucleotides from MWG (SEH4-T and SEH4-B, table 2.4) between the restriction sites Sall and BamHI of the relevant pSA1121-type plasmid (full-length Q115R,  $\Delta$ 101 or  $\Delta$ 109 Sin).



**Figure 3.13: *In vitro* binding to RTn3R Site I**

Binding of full-length WT Sin and truncated 'tadpole' constructs to the RTn3R site I (left) and binding of the same 'tadpoles' to the RSinR site (right).



**Figure 3.14: Structure of a hypothetical  $\gamma\delta$  resolvase tadpole dimer bound to DNA**

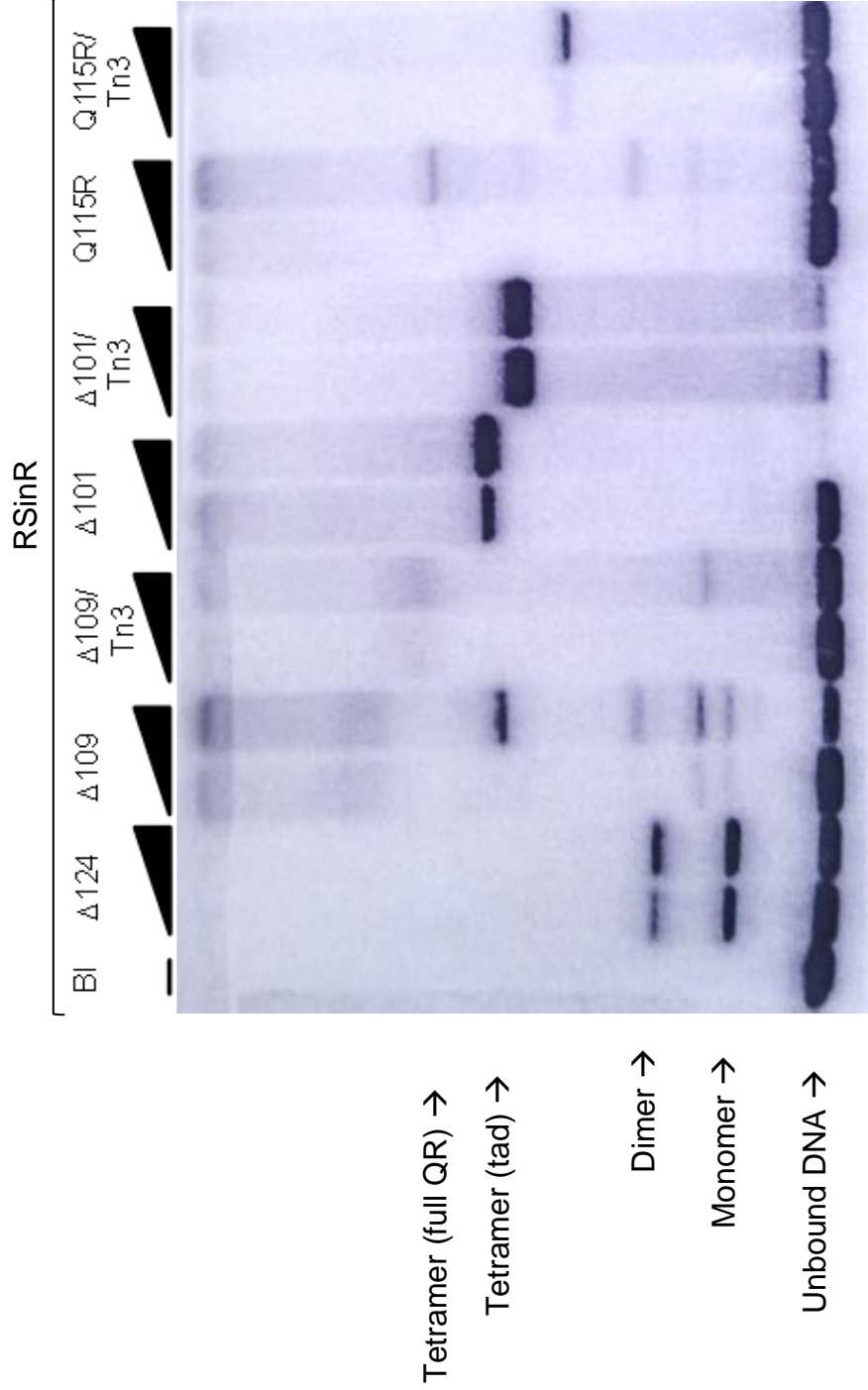
Model and cartoon illustrating a dimer of  $\gamma\delta$  resolvase bound to DNA, with  $\gamma\delta$  E-helix 'arm' region (residues 126-140) highlighted (green).  $\gamma\delta$  'arm' residues (red box) used to replace Sin residues 132-146 in 'arm'-swapping experiments. Image created using 1GDT crystal structure from PDB (Yang & Steitz 1995).



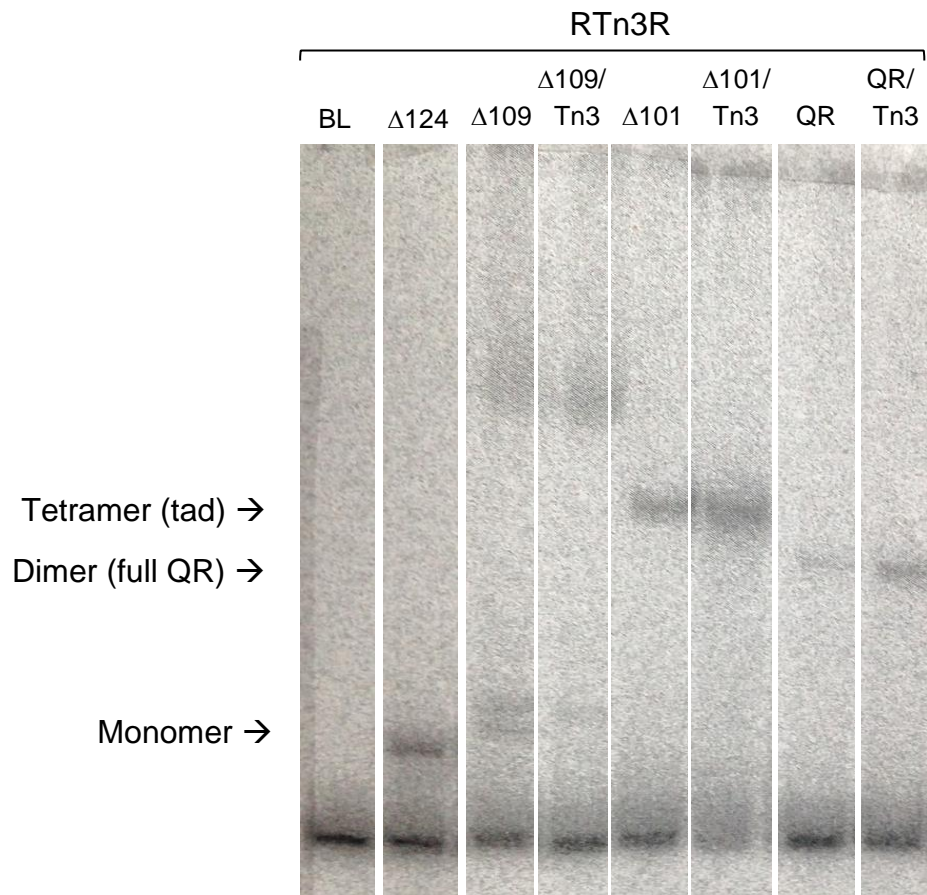
Plasmids expressing these proteins were chemically transformed (section 2.9.i) into BL21 (DE3) [pLysS] cells (table 2.1), and expressed and purified as described in section 2.22. When tested for binding capability with the RSinR substrate by polyacrylamide gel electrophoresis under the conditions previously described in section 3.3 (figure 3.15), the  $\Delta 109$  'tadpole' with the 'arm' residues 132-146 altered to the equivalent Tn3 residues ( $\Delta 109/\text{Tn3}$ ) was observed to be unable to form dimer and tetramer interactions. As predicted, alteration of the 'arm' region knocks down binding and synapsis in this hybrid. The hybrid  $\Delta 101$  'tadpole' ( $\Delta 101/\text{Tn3}$ ) appears to retain the ability to bind as a tetramer to this site; however this complex runs at a faster rate on the gel, suggesting an altered, more compact protein complex conformation.

The full-length Q115R Sin with the altered 'arm' region (Q115R/Tn3), similarly to the  $\Delta 109$  hybrid 'tadpole' was observed to lose the ability to form a tetramer with the RSinR site.

To determine whether alteration of the *res* site centre to Tn3 sequence is able to rescue this phenotype, (as may be predicted if this 'arm' region is specifically interacting with the DNA), the hybrid proteins were also tested for binding of the RTn3R substrate. This assay (figure 3.16) showed less of an effect than had been hypothesised. The hybrid full-length Q115R protein which formed dimer (but not synapse) on the RSinR substrate is also observed to form only dimer complex on the RTn3R substrate. The  $\Delta 109$  'tadpole' is similarly unaffected by the alteration of the central sequence of the site. The  $\Delta 101/\text{Tn3}$  hybrid may form a slightly more stable tetramer complex than the  $\Delta 109$  WT 'tadpole' as there is less unbound DNA observed in the hybrid than in the WT protein. The effect of altering the E-helix and the site to Tn3 sequence was not as marked as was predicted *in vitro*.



**Figure 3.15: *In vitro* binding assay of Sin ‘tadpoles’ with and without Tn3 E-helix ‘arm’ swap**  
*In vitro* binding assay showing binding of 58 bp  $\gamma^{32}\text{P}$  labelled RSinR Site I by  $\Delta 124$ ,  $\Delta 109$  and  $\Delta 101$  tadpole constructs,  $\Delta 101/\text{Tn3}$  and  $\Delta 109/\text{Tn3}$  and full-length Q115R and Full-length Q115R/Tn3 constructs with the E-helix arm residues altered from Sin to Tn3 residues (as described in section 3.2).



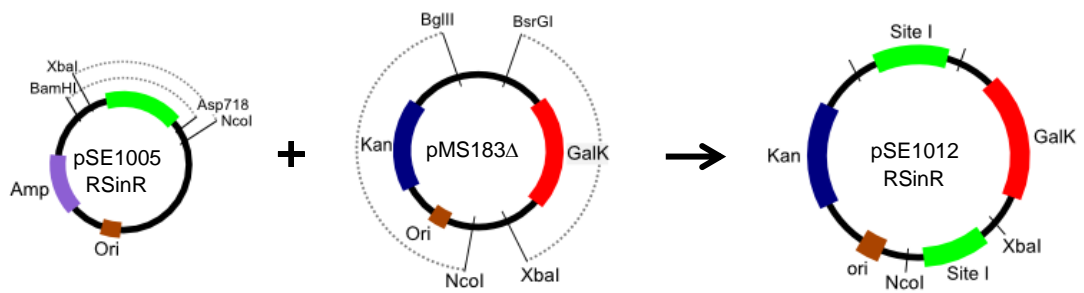
**Figure 3.16: *In vitro* binding assay of Sin ‘tadpoles’ with and without Tn3 E-helix ‘arm’ swap**

*In vitro* binding assay showing binding of 58 bp  $\gamma^{32}\text{P}$  labelled RTn3R Site I by  $\Delta 124$ ,  $\Delta 109$  and  $\Delta 101$  tadpole constructs,  $\Delta 101/\text{Tn3}$  and  $\Delta 109/\text{Tn3}$  and full-length Q115R and full-length Q115R/Tn3 constructs (QR) with the E-helix arm residues altered from Sin to Tn3 residues (as described in section 3.2).

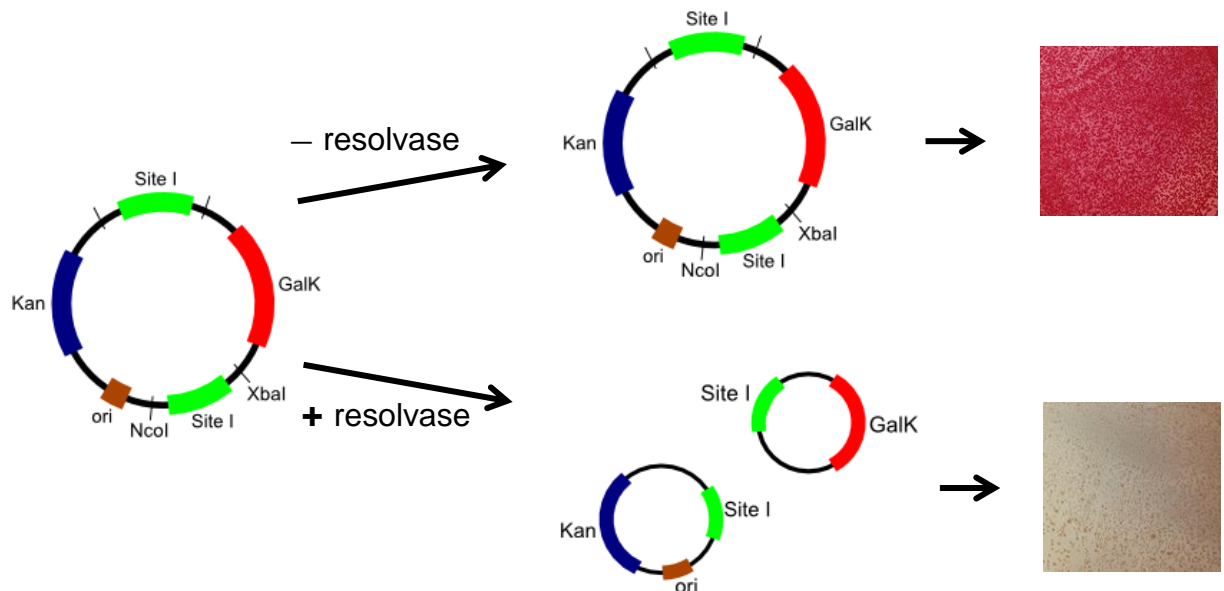
The *in vivo* effect of these alterations on the ability of the full-length proteins to carry out recombination was then assayed using MacConkey indicator plates (as described in section 2.21). *In vivo* 2-site recombination substrates were constructed by a four piece fragment swap between a pMTL23-type substrate containing a single copy of the site and pMS183 $\Delta$  (table 2.5). This construction uses restriction fragments BamHI-Asp718 and XbaI-NcoI from the pMTL23-type plasmid and BsrGI-Xba and NcoI-BglII, utilising the compatibility of BamHI/BglII and BsrGI/Asp718 (figure 3.17).

Assays were carried out *in vivo* with these substrates on MacConkey indicator plates (figure 3.18). It was observed that the R<sub>Sin</sub>R site I substrate was recombined efficiently by Q115R Sin, while the substrates with Tn3 sequence alterations at the centre (RTn3R) and the half-site ( $\Delta$ R) are not recombined. The hybrid enzymes with the 'arm' region of the E-helix replaced by the structurally equivalent residues from Tn3 resolvase, as discussed in the previous section, were also tested. The results of this assay agree with the *in vitro* observations, as the hybrid failed to recombine the RTn3R substrate (figure 3.18).

**A**

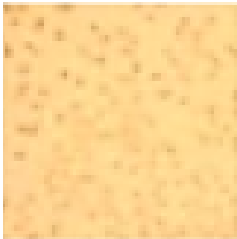
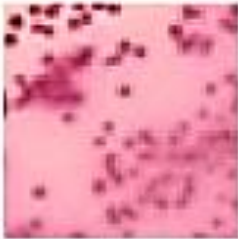



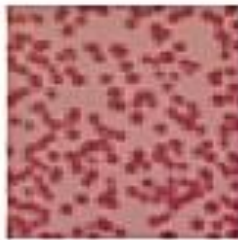


**B**



**Figure 3.17: *In vivo* recombination substrates and products of recombination**

Construction of low copy number *in vivo* substrate plasmids containing two copies of RSinR site I flanking a *galK* gene and a kanamycin resistance gene. **(A)** Plasmids were constructed by a four-piece ligation of XbaI-Asp78I + NcoI-XbaI fragments from a plasmid with a single copy of RSinR site I (pSE1005) and BsrGI-XbaI + BglII-NcoI fragments from the *galK* containing plasmid (pMS183Δ). **(B)** Outcomes in the presence and absence of resolvase. Unresolved substrate (*galK*<sup>+</sup>) produces red colonies, resolved substrate (*galK*<sup>-</sup>) produces white colonies.

	RSinR	RTn3R	$\Delta$ R
Q115R Sin			
Q115R/Tn3 Sin			

**Figure 3.18: *In vivo* recombination assay**

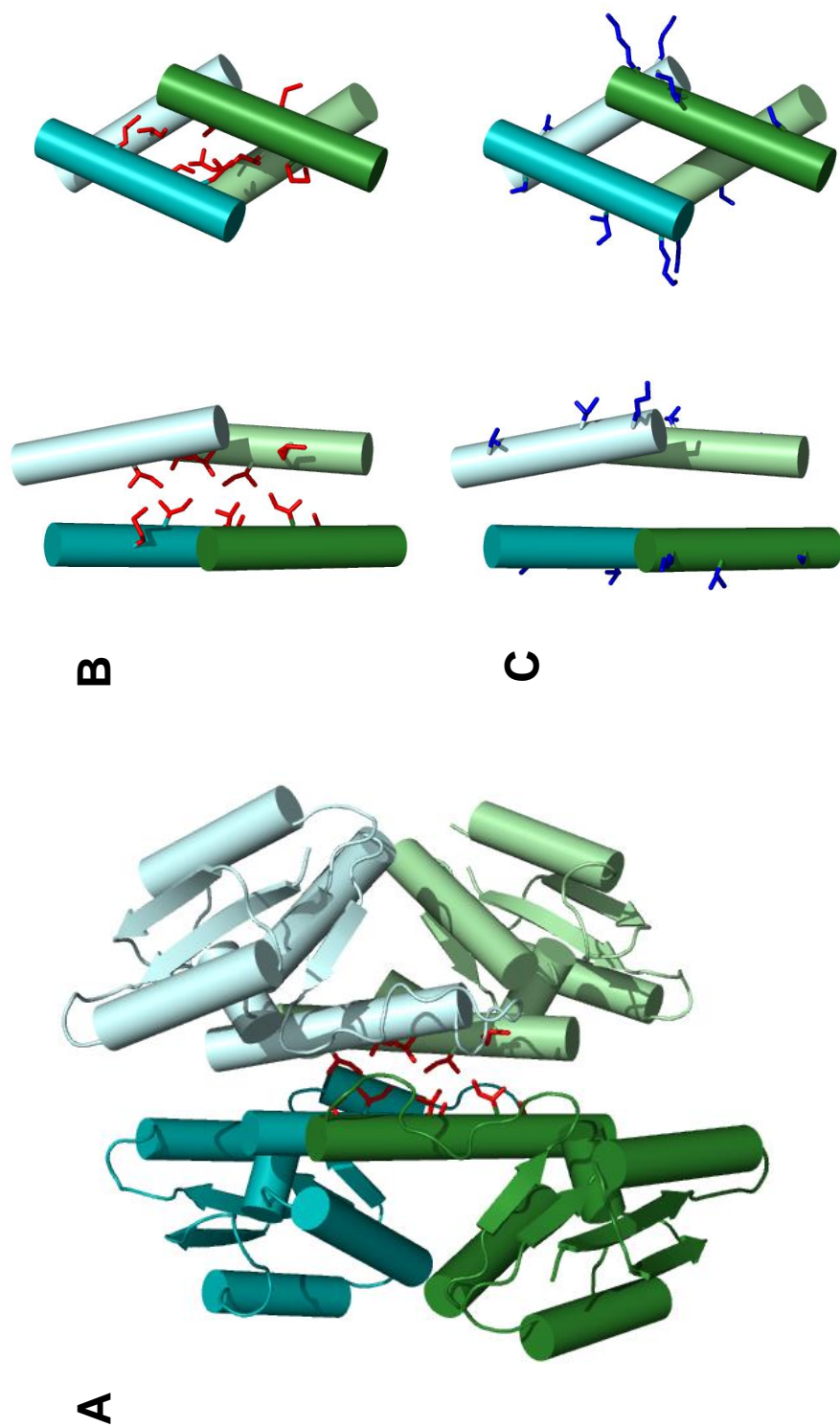
*In vivo* MacConkey assay, testing the ability of Q115R Sin to recombine the RSinR, RTn3R and  $\Delta$ R 2-site *in vivo* recombination substrates as described in section 2.7.v with a *galK* gene flanked by two *res* site Is. White colonies represent successful recombination and red colonies contain un-recombined substrate.

### 3.6 Mutations of the Sin E-helix

In order to determine which of the individual residues in the E-helix region are required for protein-protein interactions, the crystal structure of Sin was examined, and residues were selected which in the structure appear to be in positions which may have an effect on tetramer formation. Residues L112, I116 and M119 were selected as having potentially important roles in synapsis due to their location on the flat interface of the E-helices in the tetramer structure (figure 3.19A). These residues also make contacts between pairs of anti-parallel E-helices in the 'rotational dimers', an interface with importance for synapsis.

The positioning of these residues in the crystal structures suggests that they may play a role in facilitation of the subunit rotation at this interface. Residues were also selected in a structure- guided manner which do not sit on the flat tetramer interface and therefore might be expected to be less important, due to their positions in the structure, appear to be less significant for synapse formation. These residues; K110, I114 and S121 (figure 3.19B) were also mutated and the mutants tested for binding and synapsis ability.

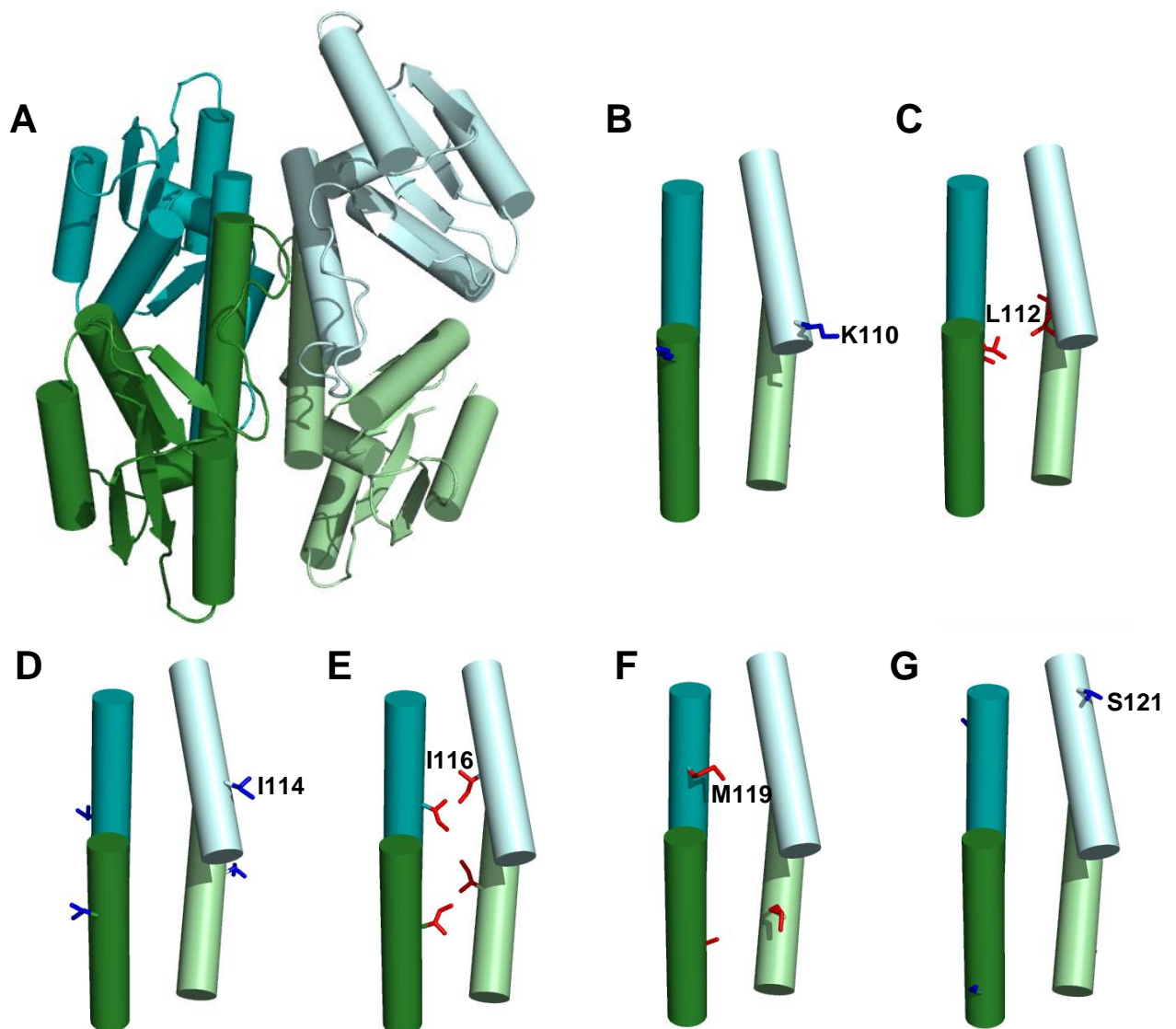
K110 is located on the outer face of the E-helix, in a position where it does not contact the E-helix of the other resolvase subunit (figure 3.20B). However it appears to have interactions with the catalytic domain. L112 is positioned on the internal surface of the E-helix, in the flat interface between the helices (figure 3.20C). I114, similarly to K110, is on the external face of the E-helix, with E-helix/catalytic domain interactions (figure 3.20D). Residues I116 and M119 are situated directly in the centre of the E-helix interface, in a position hypothesised to be important for the synapsis interface (figures 3.20E and 3.20F). The final residue selected to be mutated; S121 is in a position outside the E-helix interface, this also has E-helix/core interactions, but would hypothetically have little effect on synapsis (figure 3.20G).



**Figure 3.19: Crystal structure of the Sin E-helix**

(A) The crystal structure of Sin resolvase catalytic domains and E-helices, with residues on the interior of Sin and on the exterior of the E-helix highlighted (red and blue respectively). (B) The E-helices of Sin, with the locations of the residues on the interface between the helices highlighted in red sticks. (C) The E-helices of Sin, with the locations of the residues on the exterior of the interface between the helices highlighted in blue sticks. Figure made using 2R0Q (Mouw et al 2008).





**Figure 3.20: Crystal structure of the Sin E-helix**

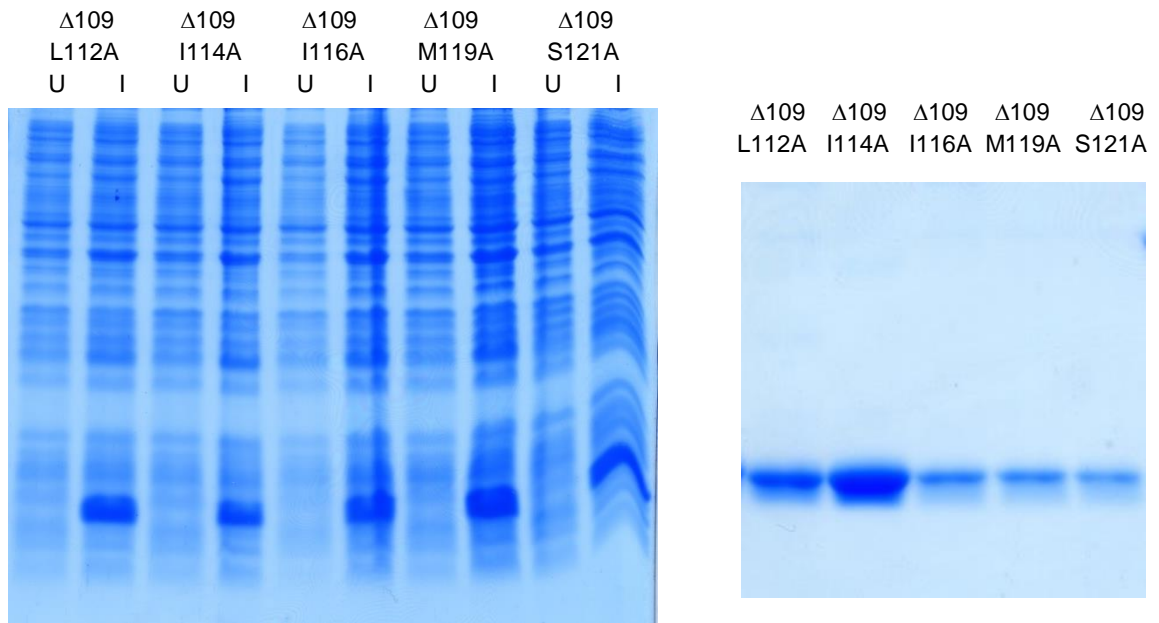
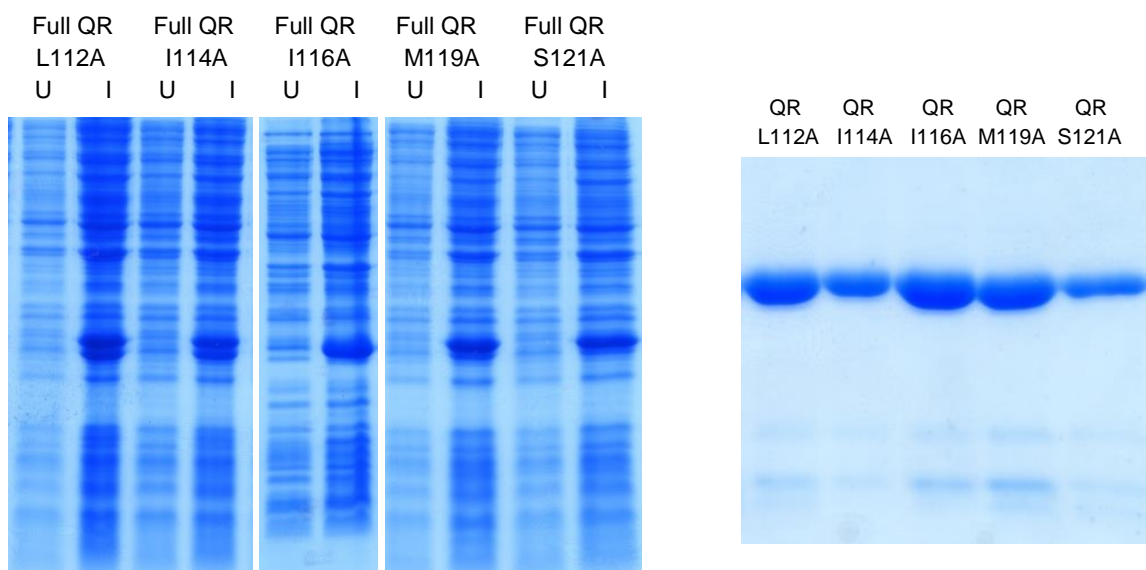
(A) The crystal structure of Sin resolvase tetramer catalytic domains and E-helices, showing the position of; (B) residue K110 on the exterior of the interface, (C) residue L112 on the interior of the interface, (D) residue I114 on the exterior of the interface, (E) residue I116 on the interior of the E-helix interface, (F) the position of residue M119 on the exterior of the interface and (G) residue S121 on the interior of the E-helix interface. Figure made using 2R0Q (Mouw et al 2008).

The chosen E-helix residues were individually mutated to alanine using synthetic oligonucleotides (SEH32-SEH45; table 2.4). Each mutation was introduced into a  $\Delta 109$  'tadpole', which was otherwise WT in sequence, and also into full-length Q115R Sin. Alanine was chosen as the residue to replace the natural residue as there is much precedence in the literature for its use in such assays (Morrison et al. 2001). Alanine has a relatively small side chain and is not disruptive to protein structure, therefore negating the chance that any observed effect is due to disruption of protein folding rather than the importance of the residue being altered.

Oligonucleotides ordered from MWG were cloned into pSA1121-type T7 inducible expression vectors already containing  $\Delta 109$  'tadpole' constructs between the restriction sites NdeI-NcoI (SEH32, 33, 34, 35; table 2.4) or NdeI-Sall (SEH36, 37; table 2.4), or vectors containing full-length Sin between BstEII-NcoI (SEH40, 41, 42, 43; table 2.4) or BstEII-Sall (SEH44, 45; table 2.4). Vectors were then cloned into BL21 (DE3) [pLysS] cells and induced, and the proteins expressed (figure 3.21) and purified as described in section 2.22.

### **3.6.i: *In vitro* binding of E-helix mutants**

Binding assays were carried out *in vitro* using  $\gamma^{32}\text{P}$ -labelled linear site I DNA constructs and purified full-length and  $\Delta 109$  tadpole versions of the E-helix mutant proteins (sections 2.25 & 2.23). These *in vitro* binding assays were designed to test two hypotheses; first, that the Sin 'tadpole' mutants are directly comparable in their behaviour and the structures which they are able to form with their full-length Q115R counterparts, and second, that the E-helix mutations selected on the basis of crystal structure data do in fact retain or reduce synapsis capability as predicted, that is, that the mutations on the crystallographic synaptic interfaces reduce or abolish synapsis, whilst those that are not on the interfaces have less of an effect or no effect.

**A****B**

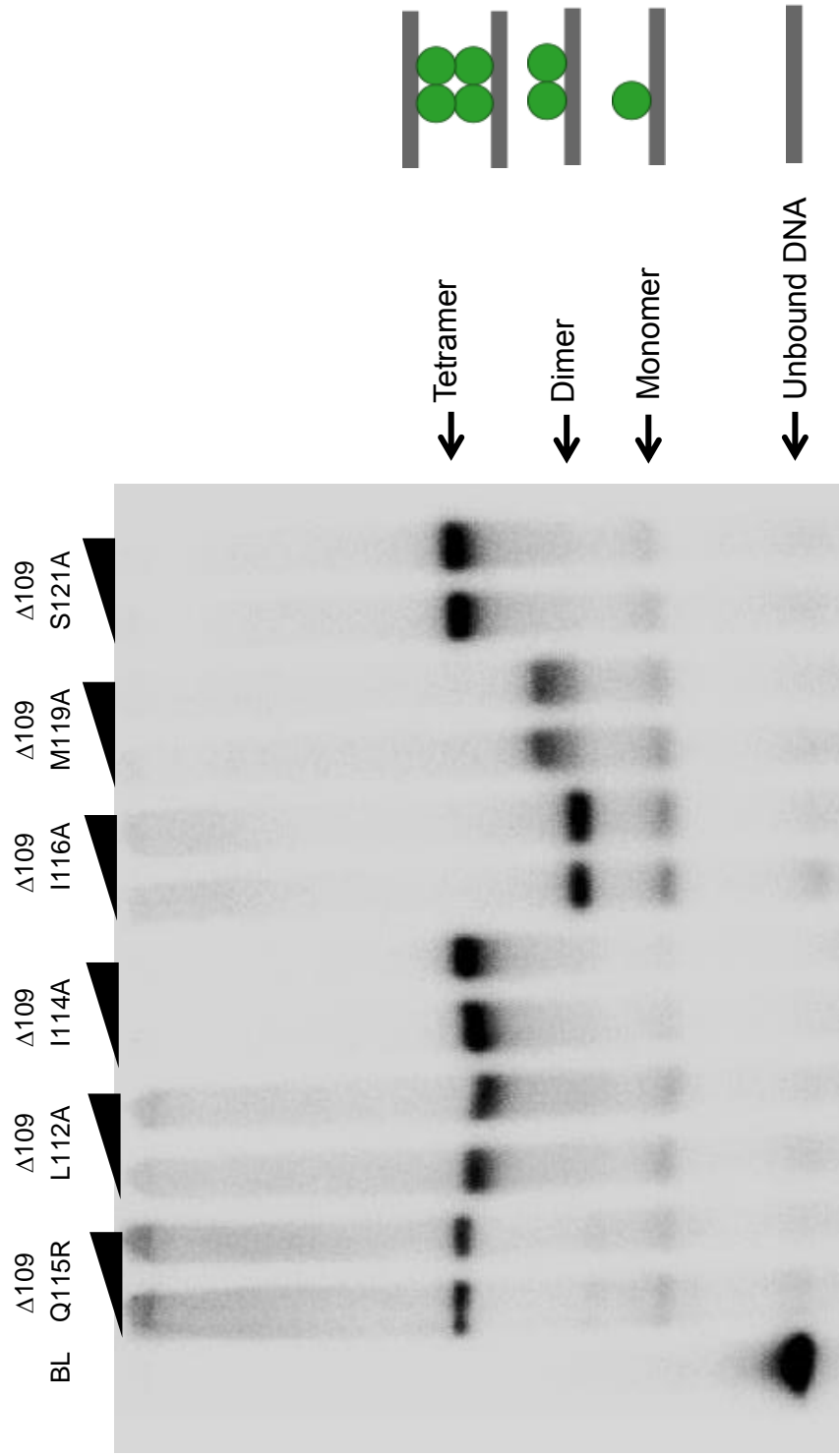
**Figure 3.21: SDS PAGE showing induced and un-induced proteins**

Polyacrylamide gel showing; **(A)** Un-induced (U) and induced (I) truncated 'tadpole' proteins with E-helix mutations (left) and the same proteins following purification (right). **(B)** Un-induced (U) and induced (I) full-length proteins with E-helix mutations (left) and the same proteins following purification (right).

The 'tadpole' mutant binding assays were carried out as described for previous binding assays. This assay (figure 3.22), showed that the  $\Delta 109/L112A$  mutant, contrary to the predicted result, retains the ability to form a tetramer complex. The observed result for the  $\Delta 109/I114A$  mutant supports the hypothesised result, with the mutation having no effect on the ability to form tetramer. Similarly, the  $\Delta 109/I116A$  mutants act as hypothesised from the crystal structure, being unable to form a synapse on the gel, showing that this residue on the E-helix interface is important for formation of the synapse. The  $\Delta 109/M119A$  mutant, which was hypothesised from the crystal structure data to be unable to form tetramer was shown to form stable dimer structures, but, as predicted, does not form a synaptic complex. The final 'tadpole' mutant tested,  $\Delta 109/S121A$ , located on the outside of the E-helix interface was also observed to agree with the predicted result, as this mutation had no effect on the ability of the protein to form a synapse.

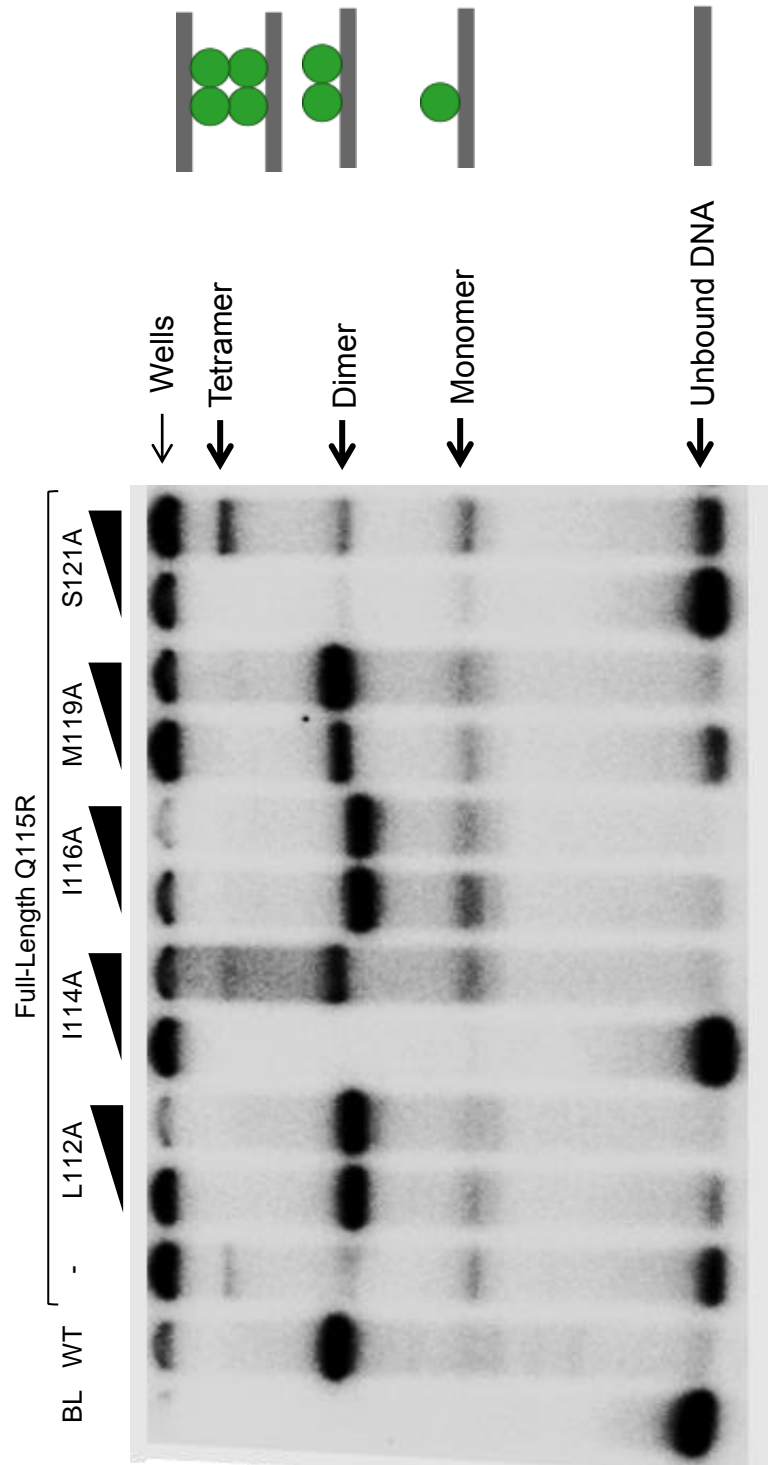
The 'tadpole' E-helix mutants were largely observed to behave as predicted from the crystal structure data. It was next desirable to test the effect of these mutations in a full-length Q115R background to determine whether, as hypothesised, the 'tadpole' constructs are analogous to full-length activated Sin.

Binding assays, again utilising the same gel conditions previously discussed were carried out. This binding assay (figure 3.23) showed that the full-length Q115R/L112A mutant, as predicted from crystal structure data, is able to form monomer and dimer complexes, but not tetramer. The full-length and  $\Delta 109/I114A$  mutant variant also supports the hypothesised result, with the mutation having no effect on the ability to form tetramer.



**Figure 3.22: *In vitro* band-shift assay**

Binding of the RSinR (58bp) res Site I by Δ109 Sin tadpoles with E-helix mutations. Unbound DNA, monomer, dimer and tetramer complexes are visible.



**Figure 3.23: *In vitro* band-shift assay**

Binding of full-length Q115R Sin + E-helix mutations (as described in section 3.3) to  $\gamma^{32}\text{P}$ -labelled RSinR site I (58bp). Unbound DNA, monomer, dimer and tetramer complexes are visible. (-) indicates Q115R Sin with no additional mutations.

The full-length Q115R/I116A and Q115R/M119A mutants, which were hypothesised to be unable to form tetramer were shown to form stable dimer structures, but, as predicted, do not form a synaptic complex. The I116A dimer complexes run at a faster rate on the gel, indicating that a more compact structure is formed by this residue change. The full-length Q115R/S121A mutant, located on the outside of the E-helix interface was also observed to agree with the predicted result, as this mutation had no effect on the ability of the protein to form a tetramer.

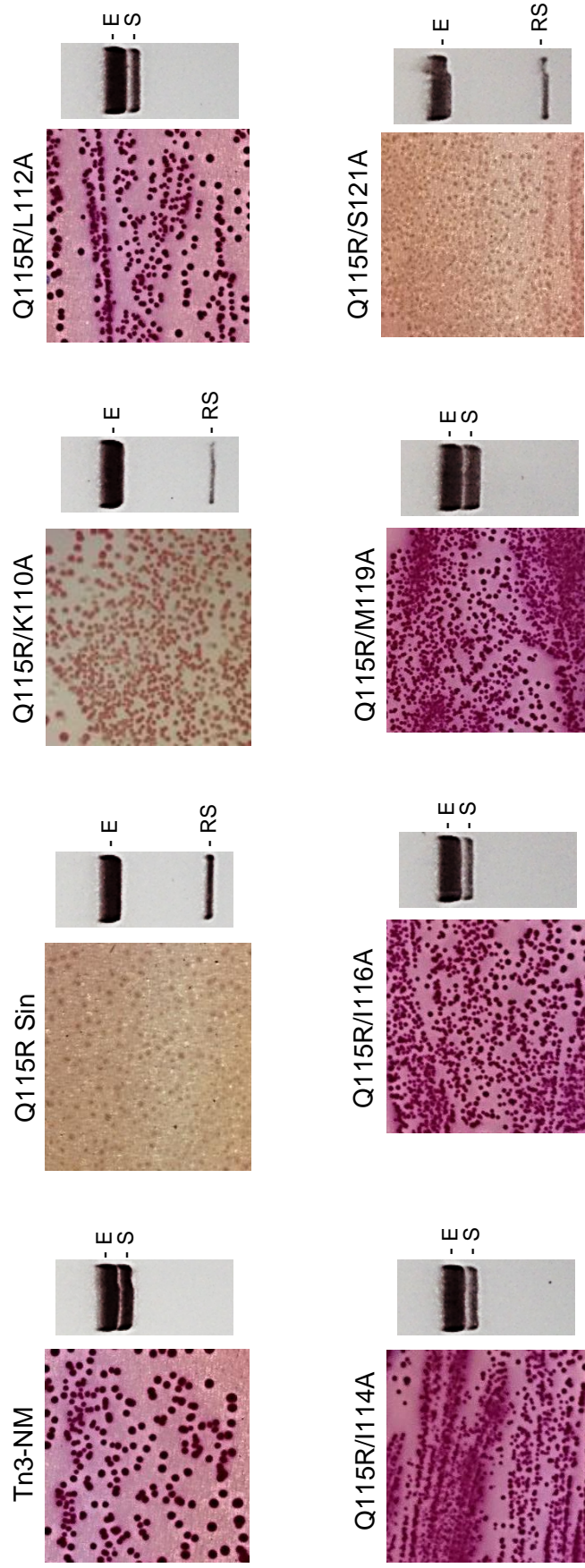
These binding assays testing ability of the ‘tadpole’ and full-length E-helix mutants to bind and form a synapse showed that all of the E-helix mutants, with the exception of the  $\Delta$ 109/L112A mutant behaved as predicted from the crystal structure data. The full-length L112A and both variants of I116A and M119A mutant, located on the E-helix interface, as hypothesised are detrimental to the formation of synapse. The  $\Delta$ 109/L112A mutant was observed to retain the ability to form a tetramer; this could be due to the position of the mutation near to the N-terminus of the truncated protein. The mutants on the outside of the E-helix interface, again as hypothesised were all observed to have no effect on the ability of the protein to form synapse.

### **3.6.ii *In vivo* recombination by E-helix mutants of Sin**

Having determined the effect of the mutations on the binding and synapsis capabilities of the ‘tadpole’ and full-length mutants, the next step was to determine the effect of the mutations on the ability of Sin to carry out recombination.

*In vivo* recombination assays utilising the 2-site RSinR substrates described in section 3.5 showed that the L112A, I114A, I116A and M119A mutations prevent catalysis of recombination by Q115R Sin (figure 3.24). The K110A mutant shows a knock-down in recombination efficiency, and S121A has no effect on the recombination ability of the protein.





**Figure 3.24: *In vivo* recombination assay E-helix mutants**

*In vivo* recombination of RSinR 2-site substrates by full-length Q115R Sin and E-helix mutants described in section 3.3. The substrate contains a *galK* gene flanked by two *res* Site I sites, recombination of these sites deletes the *galK* gene and produces white colonies, thus white colonies (panels B, C and H) represent successful recombination and red colonies (panels A, D, E, F and G) are un-recombined substrate. Tn3 NM resolvase was also tested on this substrate as a negative control. The upper band on the gel images (to the right of the indicator plates) is the expression plasmid (E) (5.5 kbp), the band directly below it is the substrate plasmid (S) (~4.9 kbp). Resolution of this substrate produces a smaller product (RS) (~3.2 kbp).



These results mostly agree with the *in vitro* binding results. The I114A mutant is able to form a tetramer, and yet is unable to catalyse recombination *in vivo*. However, perhaps it is not unexpected that the I114A mutation has detrimental effects on recombination activity due to the close proximity of this mutation to the single activating mutation (Q115R) which may be causing a slight conformational change, negatively impacting on the hyperactivity of the Q115R mutant; the WT protein is synapsis-deficient at site I in the absence of the accessory sites.

### 3.7 Discussion

In this chapter, the sequence selectivity of Sin resolvase-mediated binding and synapsis of site I was tested by assaying the effect of alteration of the nucleotide sequence at the centre of the site and by testing three truncated variants of Sin ( $\Delta 101$ ,  $\Delta 109$  and  $\Delta 124$ ) on these altered site I sequences to determine the effect of these changes on protein-protein and protein-DNA interactions. It was shown that altering the nucleotides at the centre of site I from the Sin nucleotides to the corresponding Tn3 nucleotides (RTn3R) has a negative effect on binding by Sin, showing that these nucleotides are recognised by the protein and required for strong binding of the site. The  $\Delta R$  'half-site' was also informative, as it demonstrated the specificity of binding by only a monomer of Sin.

Assays with the truncated 'tadpole' variants showed that  $\Delta 101$  and  $\Delta 109$  were competent for binding and synapsis, but deletion of the residues between 109 and 124 of the E-helix in  $\Delta 124$  prevented the formation of a synapse. This result indicated that residues in this area of the protein are required for formation of a tetramer of E-helices.

The  $\Delta 109$  and  $\Delta 124$  proteins with altered Tn3 'arm', as predicted if this region as hypothesised confers sequence selectivity, were not able to form a synapse at the R<sub>Sin</sub>R site, thus demonstrating that residues within the E-helix 'arm' region do have specificity for binding at the site. The  $\Delta 101$  Tn3 'arm' mutant however was shown to be able to form a tetramer *in vitro*, which may be due to co-operative binding and the E-helix interactions of the protein conferring stability to the protein. These mutants of Sin with alterations to Tn3 'arm' residues were also shown to be unable to form tetramer complexes or to recombine the RTn3R site. This is not as predicted, as it was hypothesised that the alteration of the nucleotides at the centre of the site might be able to rescue the ability to form tetramer complexes. This result may be due to deletion of conserved Sin residues which bind at the centre of the site, which are essential for the recombination reaction.

The E-helix mutants discussed in section 3.6 were tested in both a full-length Q115R Sin background and in a  $\Delta 109$  'tadpole' background. This was to test two hypotheses, first, that the 'tadpole' variants behave analogously to the full-length Q115R variants, based on the hypothesis that the E-helix bundle at the centre of the tetramer acts as a basic synapsis module, which is synapsis-competent, and secondly to determine whether the selected E-helix mutations have an effect on synapsis in these backgrounds.

It was determined that the 'tadpole' proteins do indeed behave in a very similar way to the full-length Q115R proteins; an interesting result, as these truncated proteins may make it easier to study binding and synapsis by resolvase free from the catalytic domains. The E-helix mutations were shown to behave largely as hypothesised, with the three mutations on the interior of the E-helix interface (L112A, I116A and M119A) preventing synapsis *in vitro* and recombination *in vivo*, and the mutations which do not sit on this interface (K110A, I114A and S121A) having no effect on binding *in vitro*.

There were however two anomalies in the behaviour of the proteins; firstly, the I114A mutant, although shown to form synapse *in vitro*, was unable to carry out recombination *in vivo*. This may be due to the proximity of this mutation to the important activating mutation Q115R, disrupting its effects. If this mutation is being negated, this would be an expected result, as the WT Sin protein is inactive in this site I x site I assay (Rowland et al. 2009).

The only differing result between the  $\Delta 109$  and full-length mutants was the L112A mutant, which was shown to be unable to form synapse in the full-length construct as hypothesised, but which was shown to retain synapsis ability in the 'tadpole'. This may be due to the location of this residue near the N-terminus of the truncated variant, which may give this protein a greater level of flexibility, allowing it to form synapse despite the mutation, by interactions of other hydrophobic residues near the N-terminus.

## Chapter 4: TALE Recombinases

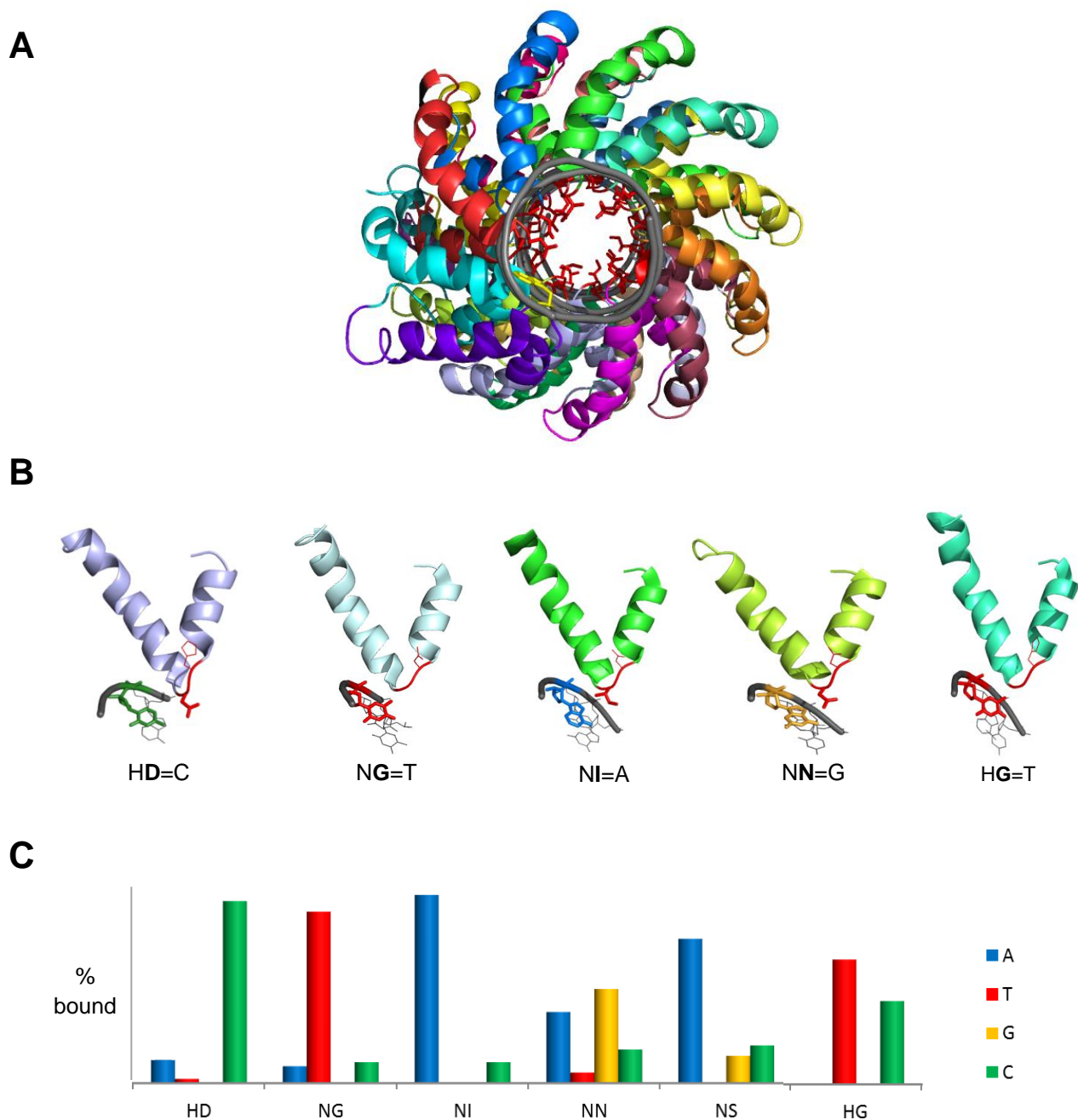
### 4.1 Introduction

The modular nature of the serine recombinases allows the DNA binding and catalytic functions of the protein to be structurally independent of each other, permitting the functional replacement of the DNA-binding domain with that of another protein. This has been demonstrated by the zinc finger recombinases, which combine the catalytic abilities of resolvase with the DNA-binding domain of a zinc finger protein, as discussed in section 1.6.iv.

This chapter describes research carried out with the aim of creating functional chimeric recombinases using a different modular DNA-binding domain, from Transcription Activator-Like Effector (TALE) proteins. This approach utilises an activated mutant of Tn3 resolvase (Tn3-NM), described in section 1.2.ii.

#### 4.1.i TALE proteins

TALEs produced by bacterial pathogens in the *Xanthomonas* sp. act naturally as transcriptional activators of virulence genes, conferring pathogenicity through specific recognition of promoter sequences (Boch et al 2010). What makes TALEs potentially very useful as tools for DNA targeting is their ability to recognise DNA specifically, utilising a central repeat domain consisting of a series of nearly identical repeats, each typically 34 amino acids in length, which usually differ only at positions 12 and 13, a region known as the repeat variable di-residue (RVD). This hyper-variable region is responsible for specific recognition of the sense strand of the DNA, with position 12 providing stability to the repeat structure while position 13 confers recognition of one base of the DNA with varying degrees of fidelity (figure 4.1), (see also section 1.6.iii).



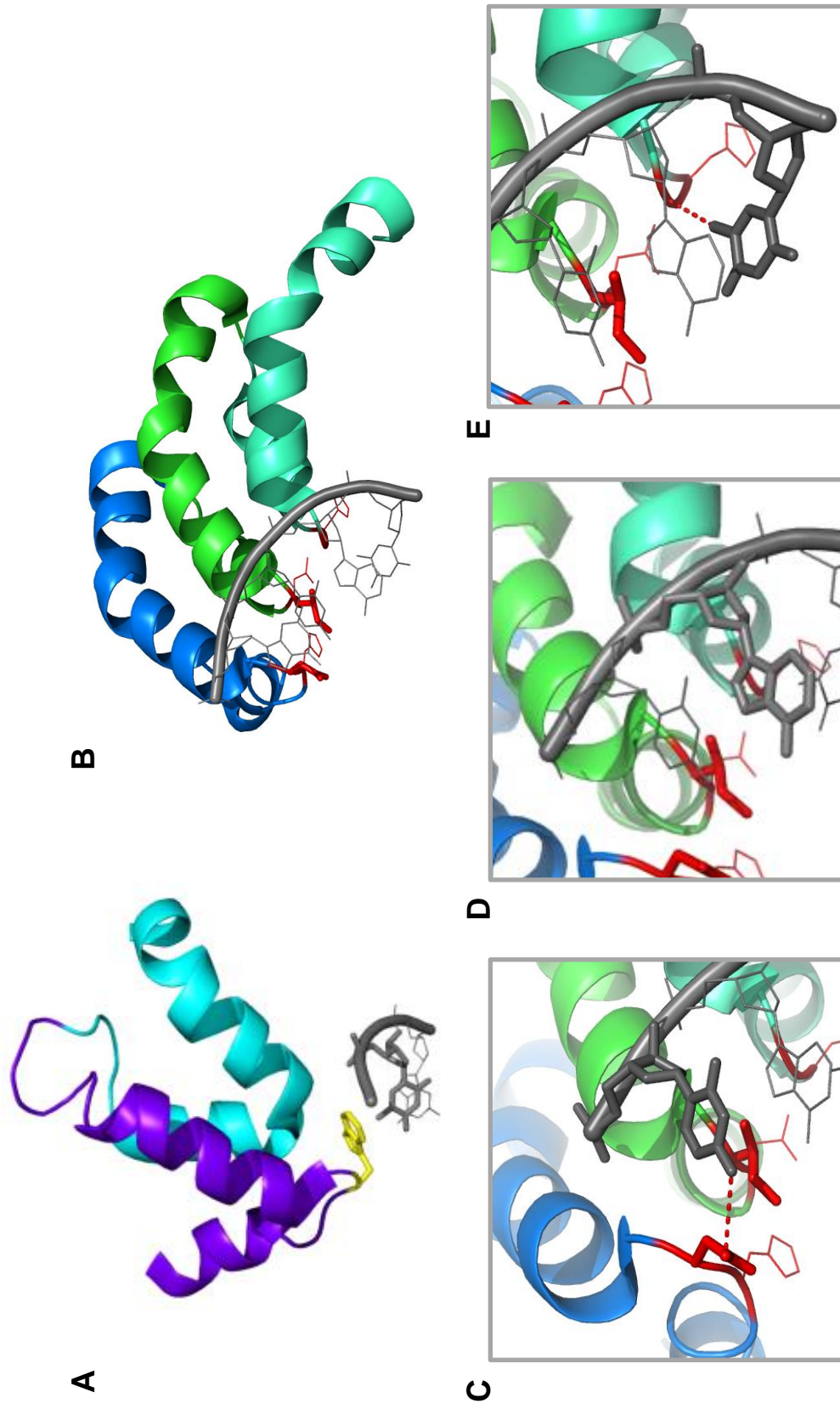
**Figure 4.1: DNA recognition by TALE RVDs**

Nucleotide recognition by the most common natural RVDs, from TALEs isolated from *Xanthomonas* sp. **(A)** Crystal structure showing the repeat fingers wrapping around the DNA, with RVDs (red sticks) positioned in the centre **(B)** Crystal structure of five common RVD at positions 12 (shown as lines) and 13 (shown as sticks), illustrating their DNA recognition (from 3UGM crystal structure (Mak et al. 2010)) **(C)** Chart created based on DNA recognition by the RVDs of the TALEs: AvrXa27, PthXo1, PthXo6, PthXo7, AvrBs3, AvrBs3Δrep16, AvrBs3Δrep109, AvrHah1 and Tal1c. (Based on RVD alignments from Moscou and Bogdanove 2009).

These repeats can be re-arranged to recognise any sequence, allowing creation of proteins with desired DNA-binding specificities. This has already been used successfully for the creation of TALENs, by fusing a FokI nuclease domain to a TALE DNA-binding domain, as discussed in section 1.6.iii.

This study employs the available crystal structures of the two TAL effectors; PthXo1 bound to DNA (pdb: 3UGM; Mak et al. 2012) and dTale2 (pdb: 4HPZ; Gao et al. 2012). Both of these models were utilised as their sequences are very similar to the TAL used in this work and are thus useful in order to design TALER constructs. The PthXo1 TALE DNA-binding domain has a 22-finger DNA recognition module with two additional imperfect repeat fingers (fingers 0 and -1) N-terminal of the RVD-containing repeats and an additional half-repeat at the C-terminus (termed finger  $\frac{1}{2}$ ) which is also involved in DNA recognition. DNA recognition by the RVDs is shown in figure 4.2. This crystal structure is disordered before residue 221 (the -1 finger in this structure beginning at residue 219), however, the more recent crystal structure of dTale2 revealed that there is structured protein N-terminal of this -1 finger, revealing a further two imperfect repeats, termed -2 and -3, which form a structure similar to the -1 finger. This crystal structure has a much smaller TALE DNA-binding domain, with only 9 RVD-containing repeats (figure 4.3). No crystal structure to date exists with structure N-terminal of this -3 finger.

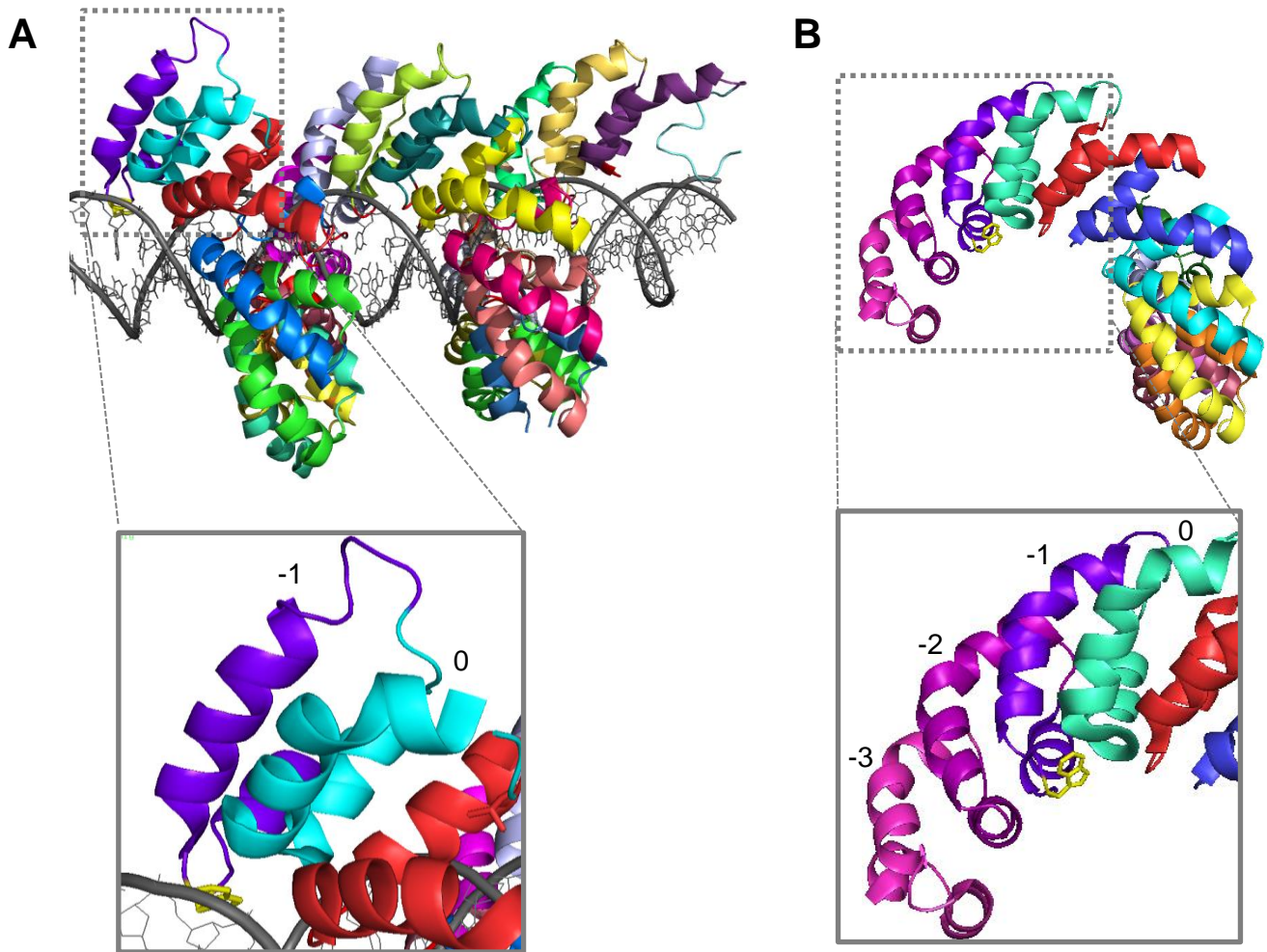
The TALE DNA-binding domain utilised in this study (*hey2*) was selected because it has previously been shown to target a TALE nuclease (TALEN) successfully to its designed site within the zebrafish genome (Sander et al 2011), so we predicted that it could also act to target a TALER to similar sequences. This TALE DBD consists of a '+63' architecture, as described by Miller et al (2011) with 63 residues C-terminal of the central repeat structure and 17 RVD containing repeats, which recognise a 17 bp DNA sequence, preceded by the conserved thymine, as discussed in section 1.6.iii (figure 4.4).



**Figure 4.2 TALE RVD-DNA interactions**

(A) Structure of N-terminal cryptic repeats, fingers -1 (purple) and 0 (cyan), showing the tryptophan residue (yellow), which forms a nonpolar van der Waals contact with the methyl carbon of the 5' thymine. (B) Structure of three TALE repeat fingers, showing their interactions with DNA. The amino acid at position 13 of each of the fingers forms specific interactions with the DNA. (C) Shows the RVD HD, which forms steric and electrostatic interactions with cytosine. (D) Shows NI which forms a desolvating interface with adenine. (E) Shows the RVD HG which forms nonpolar interactions between the glycine  $\alpha$ -carbon and the methyl group of thymine. Figure made using crystal structure 3UGM (Mak et al. 2012).





**Figure 4.3:** Structures of the DNA-binding domains of TALEs PthXo1 and dTale2

**(A)** Crystal structure of PthXo1 bound to DNA. This structure shows TALE fingers -1 (purple) and 0 (cyan), and 22 RVD containing repeat fingers coloured in rainbow. In this structure, the  $\frac{1}{2}$  repeat C-terminal of the repeats is disordered.

**(B)** Crystal structure of dTale2 (no DNA). This structure shows TALE fingers -1 (purple) and 0 (cyan) as in the PthXo1 structure, and additionally has two fingers at the N-terminus which were previously disordered in crystal structures, termed finger -3 and -2 (pink and magenta respectively). This TALE has 9 RVD containing repeat fingers coloured in rainbow. In this structure, the  $\frac{1}{2}$  repeat C-terminal of the repeats is shown to form a structure similar to the other repeats.

(Figures made using the crystal structures 3UGM (PthXo1) and 4HPZ (dTale2)).



	H	N	N	H	H	N	N	N	N	H	H	N	H	N	N	H	N
	D	G	G	D	D	N	G	G	G	D	D	I	D	I	G	D	G
t	C	T	T	C	C	G	T	T	T	C	C	A	C	A	T	C	T

**Figure 4.4:** RVD-DNA Recognition by TALE

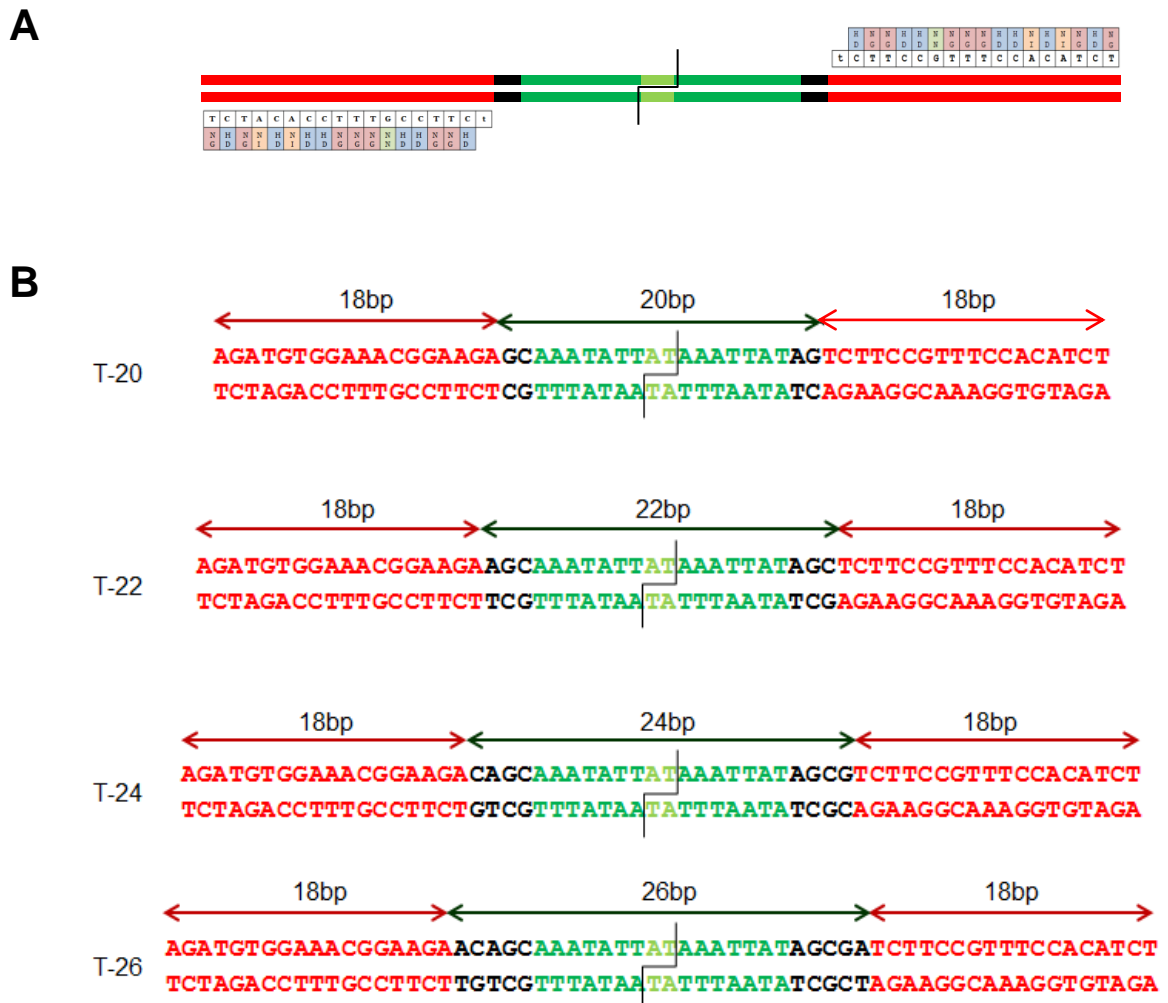
The RVDs which are present in the *hey2* TALE binding domain utilised in this study (coloured boxes) and the base which they recognise (white boxes) are shown. The required conserved 5' 'T' base is also shown.

## 4.2 Designing new TALERs

The design of TALER proteins differs from the design of the TALENs in that the TALENs have the nuclease domain attached to the C-terminus of TALE DNA-binding domain, while it is deemed necessary to attach resolvase to the N-terminus of the DNA-binding domain to create a structure analogous to that of the ZFRs which have been seen to be functional (section 1.6.iv). Indeed, the ZFRs were utilised as a basis for the break-points used for resolvase in the design of the TALER constructs to be discussed in this chapter.

The 'activated' Tn3 resolvase NM was selected as the catalytic domain donor for the creation of TALER constructs. Tn3-NM resolvase has been previously used by our group for creation of active ZFRs (Prorocic et al. 2011; Proudfoot et al. 2011), and as such, was deemed to be a good catalytic domain donor for the creation of TALERs. Additionally, some useful plasmid constructs were available for construction of TALERs.

TALERs are intended to bind to and catalyse a recombination reaction on a site consisting of TALE recognition sites flanking the central sequence of the Tn3 *res* site I. Four variants of these sites, termed 'T-sites' were constructed, with different lengths of spacer between the central sequence from the Tn3 *res* site I and the flanking TALE binding sites. These sites, named T-20, T-22, T-26 and T-28 after the number of bases separating the DNA-binding domains, are shown in figure 4.5.



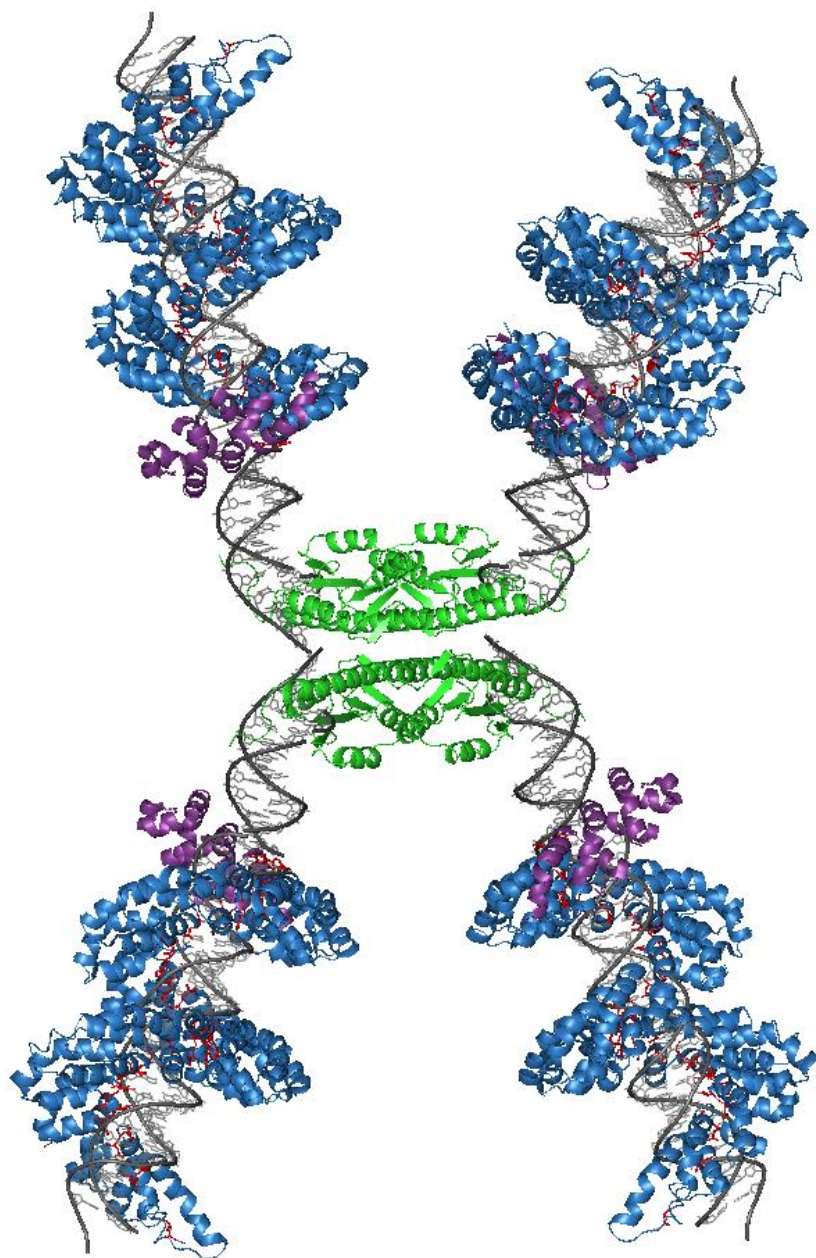
**Figure 4.5:** The T-Sites used as substrates in this study

**(A)** Cartoon depicting position of TALE binding to the sites (as in figure 4.4). The TALE DNA-binding sites, the central Tn3 *res* sequence, and the intervening spacer sequence are highlighted in red, green and black respectively. **(B)** Sites were designed with different length spacers (black) between the central 16 bp *res* site of Tn3 *res* site I (green) and the DNA-binding sites of TALE (red). Sites are named according to the number of base pairs between the DNA binding sites; T-20, T-22, T-24 and T-26.

Having selected the catalytic domain, it was next necessary to choose a breakpoint for the resolvase- that is, the position at which to detach the HTH DBD of resolvase and attach the TALE construct. The available crystal structure data from the closely related  $\gamma\delta$  resolvase (Yang and Steitz 1995, Li et al 2005), and the position utilised in the zinc finger resolvase work previously carried out in the Stark group (Prorocic et al 2011) provided a logical position for this attachment; retaining resolvase sequence up to residue 148. A model of a potential TALER construct is shown in figure 4.6.

The next step in creating a functional chimeric protein was to select the appropriate linker length for attachment of the Tn3 catalytic domain to the DNA-binding domain of the TALE and to determine how much of the TAL N-terminus is required for optimal activity.

Previous studies involving TALEs have indicated that the truncation of the TALE domain's N-terminus can sometimes improve the protein's activity (Miller et al 2010, Zhang et al 2011, Mercer et al 2013). Therefore, in order to determine what length of TALE domain is optimal for TALER constructs, several N-terminal truncations ( $\Delta 48$ ,  $\Delta 84$ ,  $\Delta 119$  and  $\Delta 149$ ) were created, and the different length TALERs were tested for binding and synapsis *in vitro* (figures 4.7 and 4.8). The  $\Delta 149$  variant begins at the start of the '-3' TALE finger. The longer truncated variants were created by addition of between 30 and 35 amino acids, to create TALE 'finger' length additions, hypothesising the presence of further 'finger' like structures N-terminal of the known structure.



**Figure 4.6:** Model of hypothetical TALER construct

This TALER model shows a potential structure formed by this chimeric protein. Model shows  $\gamma\delta$  resolvase fused to TALE residues from 154 (finger -1) to the  $\frac{1}{2}$  repeat finger at the C-terminal end of the TALER binding domain.

Model constructed by superimposing crystal structures of  $\gamma\delta$  resolvase (residues 1-144, green) comprising the catalytic NTD and E-helix (1ZR2, Li et al 2005), dTale2 (residues 154-254, purple), with fingers -3, -2 and -1 (4HPZ, Gao et al 2012) and PthXo1 (residues 255-1048, blue), containing finger 0, the 22 RVD containing repeats and a final  $\frac{1}{2}$  repeat (3UGM, Mak et al 2012). RVDs are highlighted in red.

[illegible]

**Figure 4.7: Sequence of the TALE (hey2) utilised in the construction of the TALERs used in this study**

Truncation points of the TALEs ( $\Delta 48$ ,  $\Delta 94$ ,  $\Delta 119$  and  $\Delta 149$ ) are highlighted (yellow), fingers -3, -2, -1 and 0 N-terminal to the repeats are highlighted (blue and purple) and repeat fingers are alternately highlighted in orange/red. The RVDs conferring DNA recognition are underlined. All TALER constructs made in this work have 63 amino acids of hey2 (green) following the final  $\frac{1}{2}$  repeat (brown) and a C-terminal histidine tag.





### 4.3 Construction of TALERs

The TALER constructs utilised in this study employ a previously used TALE DNA binding domain from a TALEN construct targeting the zebra fish *hey2* gene (TALEN1297, Sander et al 2011). This construct has a  $\Delta 222$  structure, with the N-terminal 222 amino acids of the TALE binding domain replaced by a 3xFLAG tag and linker sequence and a C-terminal Fok1 nuclease domain attachment. In order to utilise this domain for TALER construction, the N-terminal TALE sequence had to be inserted and the C-terminal nuclease domain removed. Due to the repeat nature of the TALE DNA-binding domains, TALE domains are difficult to synthesise, thus, two synthetic constructs were designed for construction of the desired TALER (figures 4.9 and 4.10). These constructs were synthesised by GeneArt.

The first of these constructs (GeneArt\_SpeI-ApaI) was designed to link the Tn3-NM catalytic domain to the N-terminus of the TALE domain, utilising the SpeI restriction site which is at the previously designed attachment point for chimeric resolvases (Akopian et al 2003). This construct was also designed with silent restriction sites at appropriate intervals to facilitate creation of the truncated TALER variants described in the previous section. The second of the GeneArt constructs (GeneArt\_MfeI-KpnI) was designed to introduce a His<sub>6</sub> tag for protein purification and termination codons at the C-terminus of the protein. These synthetic constructs were cloned into a high copy number T7 inducible expression vector with the TALEN-1297 repeat sequence in order to create the full-length TALER construct (figure 4.11). The nucleotide sequences ordered from MWG are shown (supplementary figures).

These expression vectors were then transformed into BL21(DE3)[pLysS] cells and protein production was initiated by IPTG induction as described in section 2.22. Following induction, samples were loaded onto an SDS polyacrylamide gel to determine whether the proteins had been expressed (figure 4.12). Following successful expression, the proteins were purified using the C-terminal His<sub>6</sub> tag and dialysed for concentration and further purification (section 2.22).

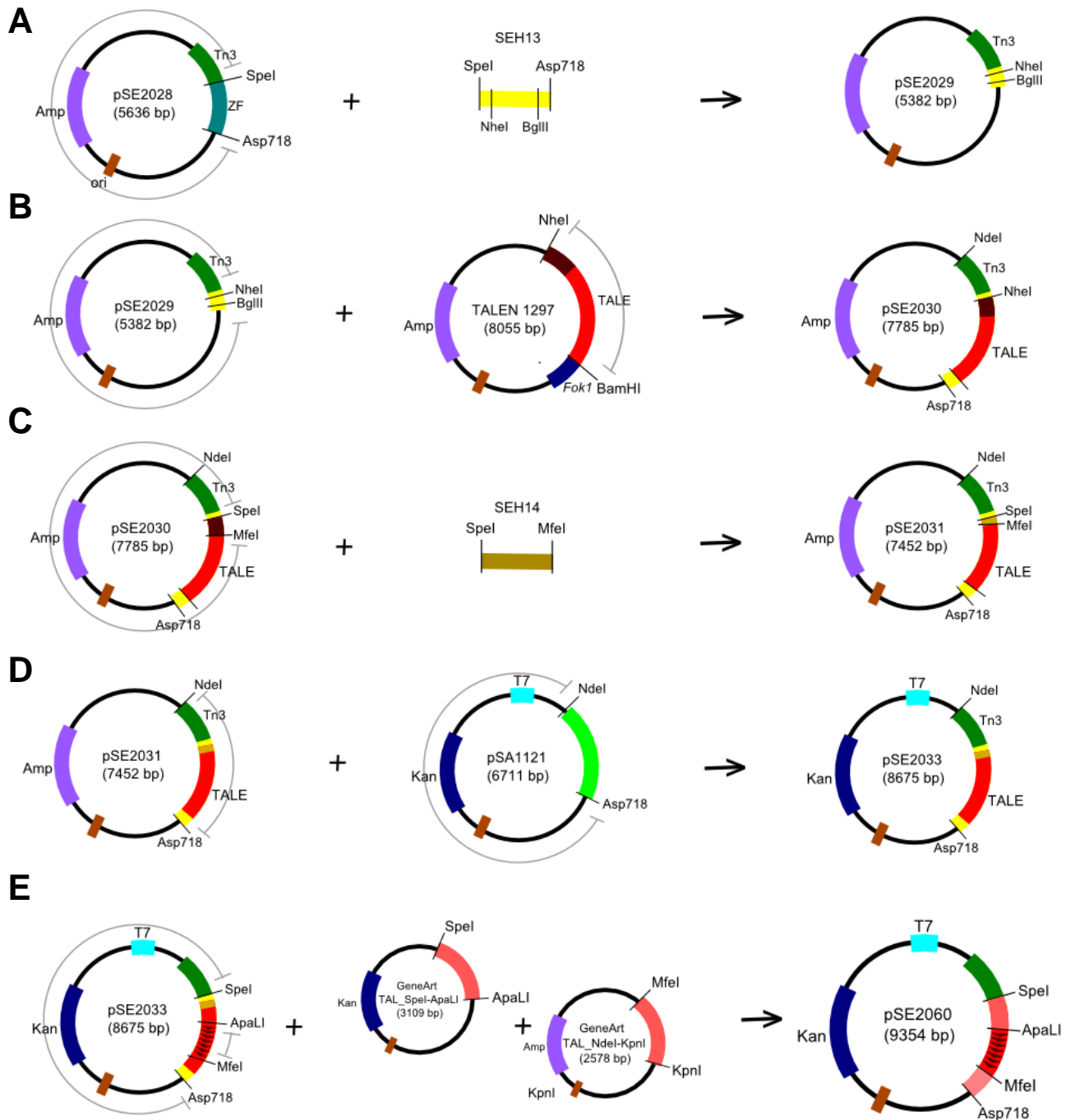






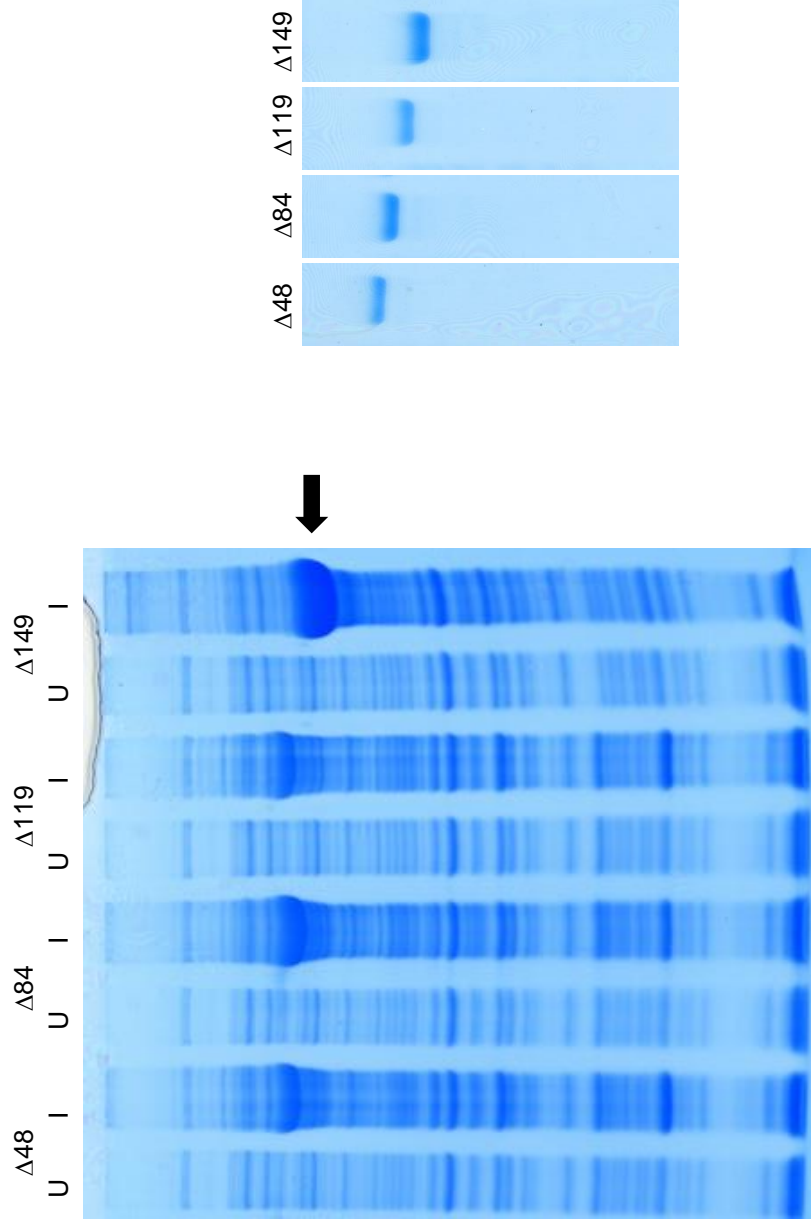
**Figure 4.10** : Synthetic GeneArt construct for creation of His-tagged Tn3-TALERS

This construct contains residues (+4 - +63) of TALE followed by a His<sub>6</sub> tag and a termination signal. Designed to be cloned on a Mfel-KpnI fragment this was used to create Tn3-TALER constructs. The sequence in red shows the designed amino acid sequence, the GeneArt vector is shown in black.



**Figure 4.11: Construction of TALER expression vectors**

Construction process for creation of expression vector containing full-length TALE (1-(+63)). **(A)** Oligonucleotide SEH13 is cloned in between SpeI and Asp718 of a ZFR containing vector to remove the zinc finger DBD and insert the restriction sites NheI and BglII. **(B)** TALER repeats from TALEN1297 (Sander et al 2011) are cloned into the Tn3 vector, using the compatibility of BamHI/BglII. **(C)** Oligonucleotide SEH14 was cloned in between SpeI and MfeI to remove a 3xFLAG tag which was present in the TALEN construct. **(D)** The TALER construct (the TAL domain starting at residue 222; TALE finger -1) was excised and inserted into a high copy number, IPTG inducible plasmid on an NdeI-Asp718 fragment. **(E)** Expression vector containing full-length TALE (1-(+63)) was constructed by fragment swap between GeneArt synthetic constructs and the  $\Delta 222$  TALER plasmid pSE2033 (position of RVD containing repeats indicated by brown arrows).



**Figure 4.12 : Polyacrylamide gel showing TALER proteins**

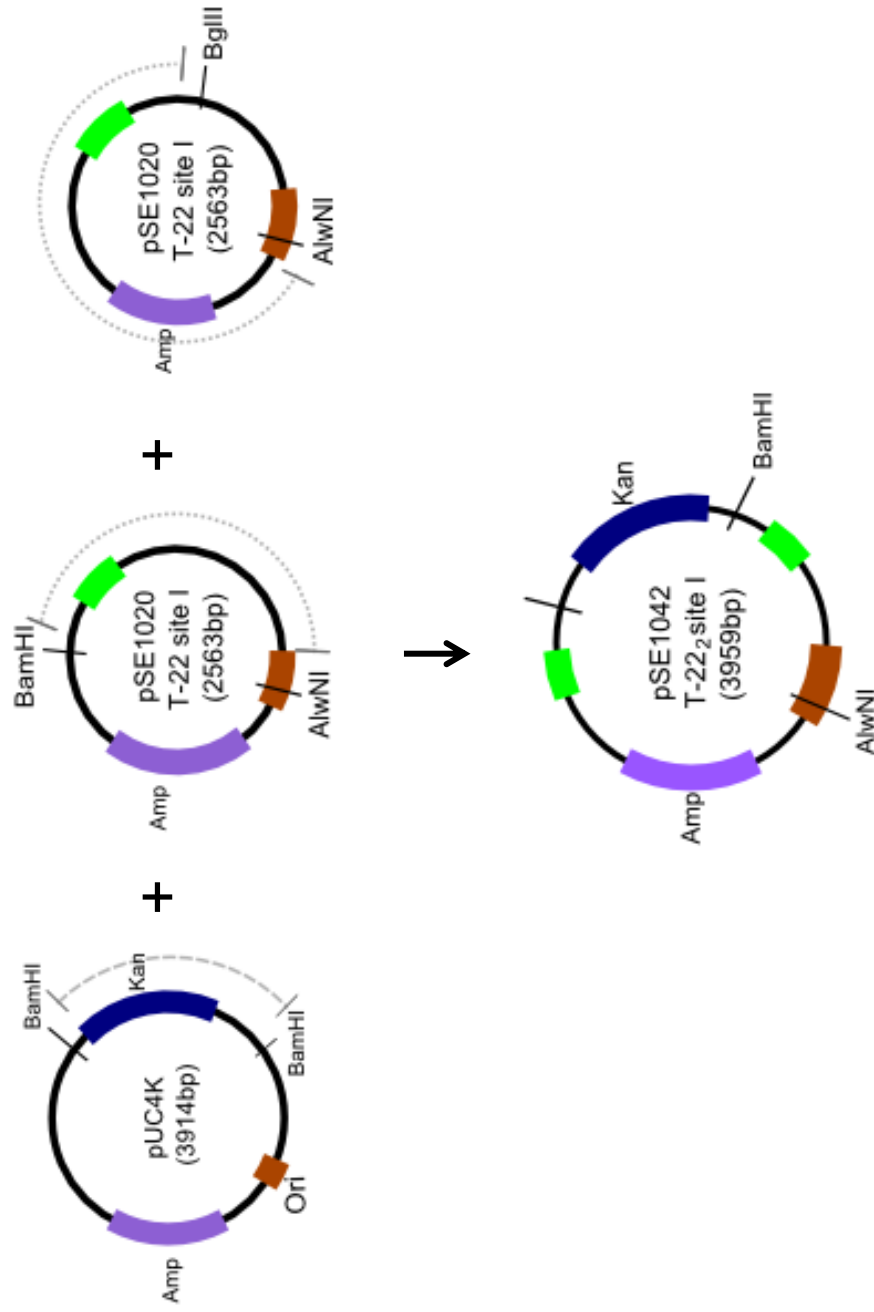
Samples were taken pre- (U) and post- (I) induction to determine whether induction was successful and run on an SDS gel. The induced TALER proteins are indicated (arrow) (left). The same protein samples following purification are shown (right).  $\Delta 119$  appears to run at a reduced rate in the induced sample compared to the purified, this may be due to gel imperfection in this lane.

#### 4.4: TALER Activity *in vivo* and *in vitro*:

TALER constructs were tested *in vivo* for recombination ability using a MacConkey assay with a two site substrate as described in section 2.21. However, these assays did not show any evidence of recombination by these proteins. All colonies produced were red in colour and making DNA from these colonies showed no evidence of recombination activity. This does not necessarily indicate that there was no activity, and may instead mean that there was a very low level of activity. *In vitro* recombination assays also showed no evidence of activity. As such, more sensitive *in vitro* cleavage assays were carried out.

*In vitro* cleavage assays utilising a high copy number 2-site substrate (figure 4.13) were carried out to test each of the four purified truncated TALER constructs,  $\Delta 48$ ,  $\Delta 84$ ,  $\Delta 119$  and  $\Delta 149$  for ability to cleave the T-sites; T-20, T-22, T-24 and T-26, to test which substrate spacing is optimal for activity of these proteins. The cleavage assay is a sensitive assay which looks at the cleavage state of the DNA directly. Cleavage at one site by the TALER would be expected to yield a single linear product, while cleavage at both of the sites should yield two linear substrates of different sizes.

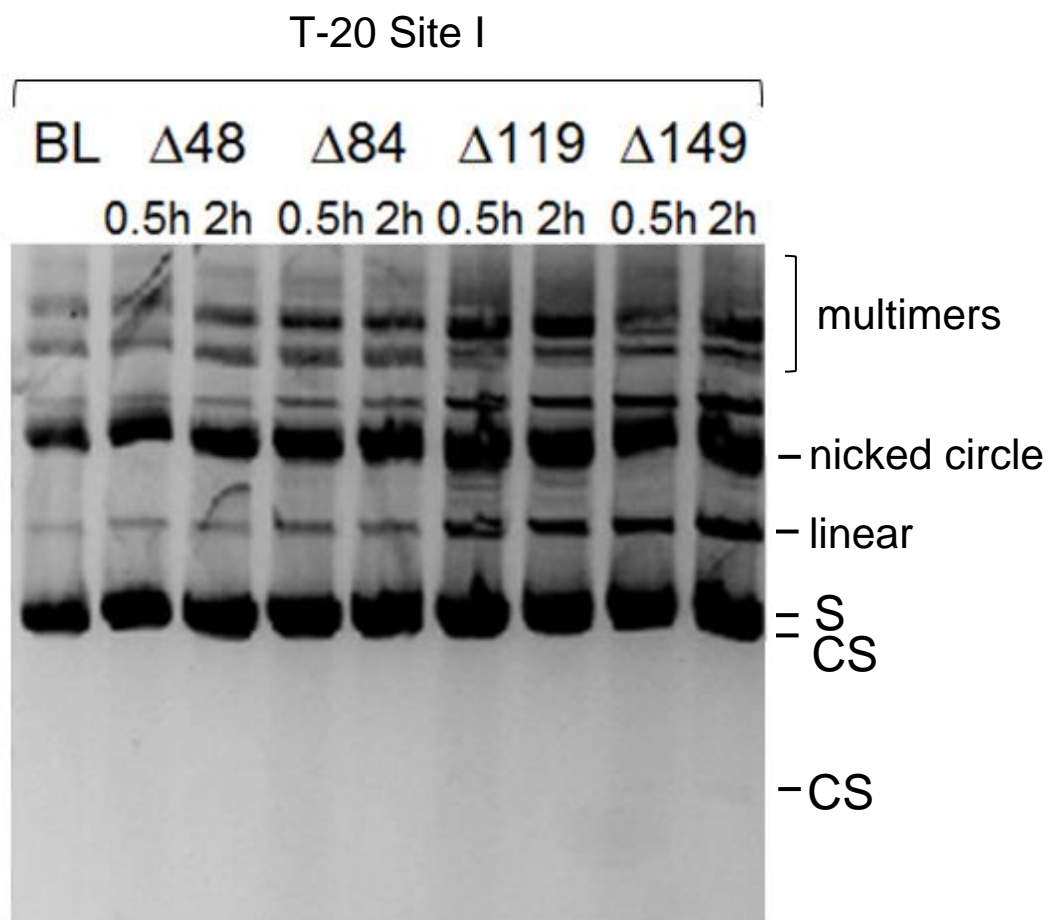
The shortest of the T-site substrates tested (T-20) was not observed to be cleaved by any of the truncated TALERS tested (figure 4.14) indicating that this substrate does not have the correct spacing to be functional. The T-22 substrate when tested was shown to be a functional substrate for cleavage by the  $\Delta 149$  TALER, though not the other three truncated variants (figure 4.15). The T-24 substrate behaved similarly to the T-22, being cleaved by the  $\Delta 149$ , but not by the other variants (figure 4.16). The longest of the substrates tested in these assays, T-26 was also a functional site for  $\Delta 149$ , although it does not appear to be as efficient as T-22 and T-24 (figure 4.17). Increased multimer formation seen with  $\Delta 149$  constructs indicates cleavage activity and increased recombination within these substrates. Multimer formation



**Figure 4.13:** Construction of high copy number *in vitro* T-site I substrates

T-sites from pMTL23-type substrates containing a single copy of the site were cloned into the pUC4K cloning vector. Sites were extracted from the single site plasmid pSE1020 on AlwNI-BamHI and AlwNI-BglII fragments and ligated with BamHI-BamHI fragment from pUC4k, utilising the compatibility of the restriction enzymes BamHI and BglII.

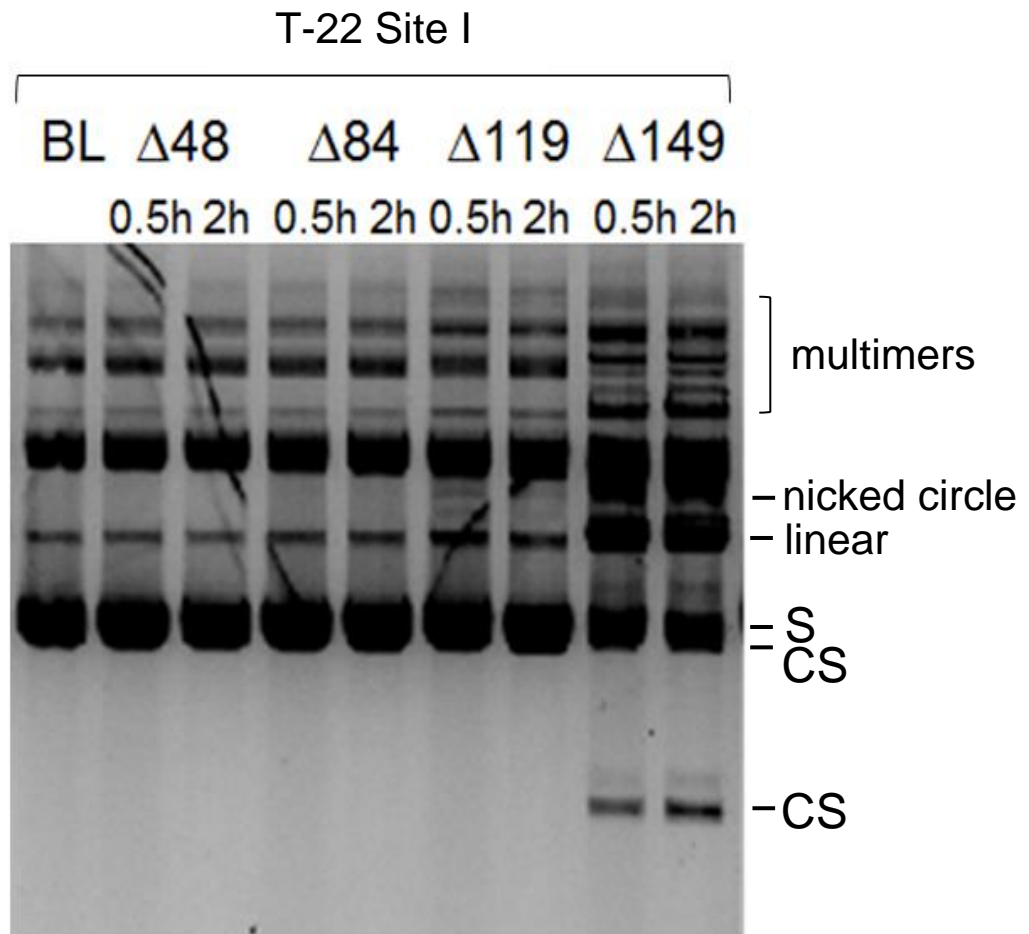
AGATGTGGAAACGGAAGAGCAATATTATAAATTATAGTCTTCCGTTTCCACATCT  
 TCTAGACCTTTGCCTTCTCGTTTATAATTTTAATATCAGAAGGCAAAGGTGTAGA



**Figure 4.14:** *In vitro* cleavage assay (T-20 site I)

Assay for cleavage of T-20 site by the  $\Delta 48$ ,  $\Delta 84$ ,  $\Delta 119$  and  $\Delta 149$  TALERs. This substrate was not acted on by any of the TALER variants with which it was tested, leaving the substrate (S) un-cleaved. Expected positions of cleaved substrate (CS) are shown. The larger cleaved product ran at a position similar to the un-cleaved supercoiled plasmid.

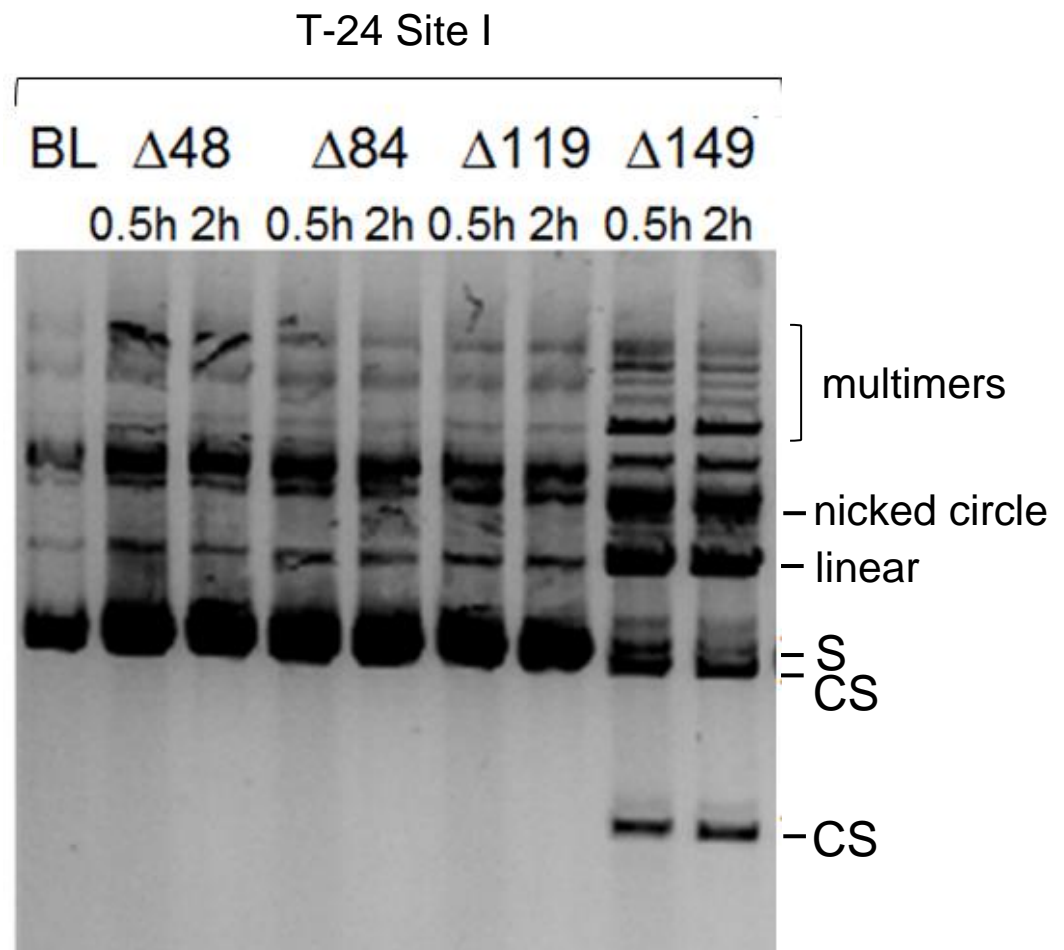
AGATGTGGAAACGGAAGAAGCAATATTATAAATTATAGCTCTTCCGTTTCCACATCT  
 TCTAGACCTTTGCCTTCTTCGTTTATAATATTTAATATCGAGAAGGCAAAGGTGTAGA



**Figure 4.15:** *In vitro* cleavage assay (T-22 site I)

Assay for cleavage of T-22 site by the  $\Delta 48$ ,  $\Delta 84$ ,  $\Delta 119$  and  $\Delta 149$  TALERs. This substrate was cleaved by the shortest of the truncated TALERs ( $\Delta 149$ ) cleaving the substrate (S) to produce smaller, cleaved products (CS). Cleavage was not seen with the longer TALERs. The larger cleaved product runs at a similar position to the un-cleaved supercoiled plasmid.

AGATGTGGAACGGAAGACAGCAAAATATTATAAATTATAGCGTCTTCCGTTTCCACATCT  
TCTAGACCTTTGCCTTCTGTCGTTTATAAATTTAATATCGCAGAAGGCAAAGGTGTAGA

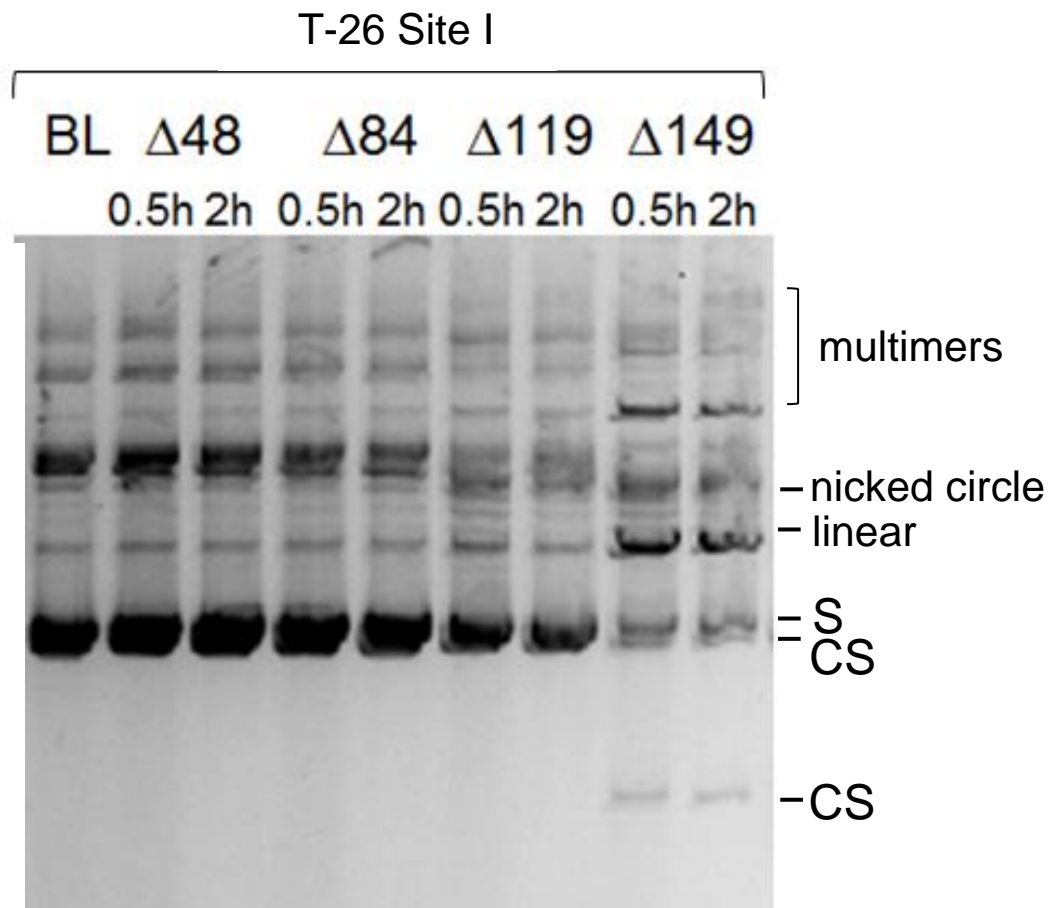


**Figure 4.16:** *In vitro* cleavage assay (T-24 site I)

Assay for cleavage of T-24 site by the  $\Delta 48$ ,  $\Delta 84$ ,  $\Delta 119$  and  $\Delta 149$  TALERs. This substrate was cleaved by the shortest of the truncated TALERs ( $\Delta 149$ ) cleaving the substrate (S) to produce smaller, cleaved products (CS). Cleavage was not seen with  $\Delta 48$ ,  $\Delta 84$  and  $\Delta 119$  TALERs. The larger cleaved product runs at a position similar to the un-cleaved supercoiled plasmid.



AGATGTGGAACGGAAGAACAGCAAAATATTATAAATTATAGCGATCTTCCGTTTCCACATCT  
TCTAGACCTTTGCCTTCTGTTCGTTTATAATATTTAATAATCGCTAGAAGGCAAAGGTGTAGA



**Figure 4.17:** *In vitro* cleavage assay (T-26 site I):

Assay for cleavage of T-26 site by the  $\Delta 48$ ,  $\Delta 84$ ,  $\Delta 119$  and  $\Delta 149$  TALERs. Cleavage was carried out only by the shortest of the truncated TALERs ( $\Delta 149$ ) acting on the substrate (S) to create smaller, cleaved products (CS). Cleavage was not seen with the longer TALER variants.

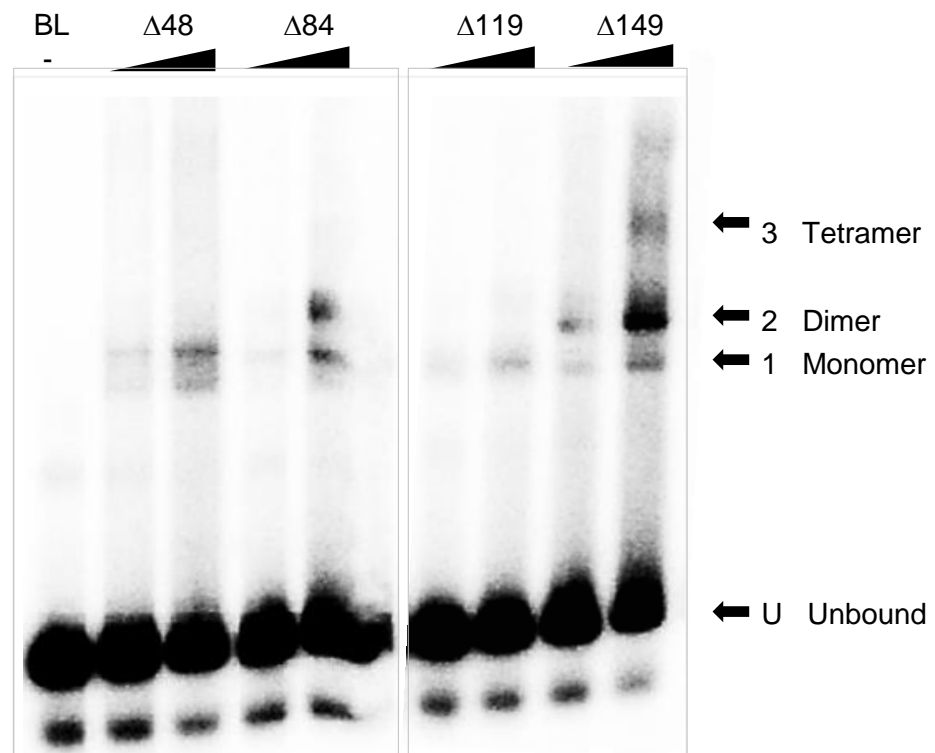
increases in the T-24 and T-26 substrates, indicating that these are being acted on more efficiently by the  $\Delta 149$  TALER.

These assays indicated that the best of the substrates tested were the T-22 and T-24 substrates which show similar amounts of cleavage products, and additionally showed that the longer truncated variants,  $\Delta 48$ ,  $\Delta 84$  and  $\Delta 119$  are not functional under these conditions, while the shortest of the TALER constructs, the  $\Delta 149$  construct, was functional on three of the four substrates. Following these *in vitro* cleavage assays therefore, the T-22 site was selected for further study.

Of the TALERs tested, only the protein with the largest TALE domain N-terminal truncation was shown to have activity on the T-site substrates. This could be because the other proteins are unable to bind to their sites, because the additional amino acids at the N-terminus of the longer constructs are folding in such a way that they are blocking protein-protein interactions and thus preventing the formation of the dimer and tetramer constructs which are necessary for activity of the resolvase catalytic domain, or because the additional length at the N-terminus is blocking access of the resolvase to the centre of the *res* site.

In order to test whether the TALER constructs are able to bind to the DNA and to determine what types of protein interactions are formed, 75 bp T-22 linear DNA ordered from MWG as oligonucleotides (SEH 48-T and SEH 48-B) was labelled with  $\gamma^{32}\text{P}$ -ATP and *in vitro* binding assays were carried out. The assays showed that the  $\Delta 48$ ,  $\Delta 84$ ,  $\Delta 119$  and  $\Delta 149$  truncated TALER variants are all able to bind to the site, indicating that the binding domains are targeting their designed sites (figure 4.18).

The results of the *in vitro* band-shift assay shown in figure 4.18 agree with the results obtained from the *in vitro* cleavage assays in that they show that the  $\Delta 149$  TALER forms three complexes on the gel, where the  $\Delta 48$ ,  $\Delta 84$  and  $\Delta 119$  constructs form only two complexes on the gel. These complexes, if the



**Figure 4.18:** *In vitro* binding band-shift assay

T-22  $\gamma^{32}\text{P}$ -ATP labelled sites bound by  $\Delta 48$ ,  $\Delta 84$ ,  $\Delta 119$  and  $\Delta 149$  TALERs. The  $\Delta 149$  construct is able to form a large complex (3) (hypothesised to be a synaptic tetramer complex) the  $\Delta 84$  and  $\Delta 149$  constructs form putative dimer structures (2) and the  $\Delta 48$  and  $\Delta 119$  structures appear to form only a hypothesised monomeric structure (1).

TALER is behaving as hypothesised, in a manner similar to the resolvases, may be representative of complexes containing one, two and four TALER subunits, similar to the complexes which were demonstrated in chapter 3 for the full-length and truncated Sin constructs. It is hypothesised that the largest of these complexes could be a synapse containing 4 protein subunits and two DNA sites. Confirming this hypothesis however requires more research to be carried out, such as the mixed-length (45 bp and 77 bp) DNA substrates used to identify the 'tadpole' synapse in chapter 3.

#### 4.5: Discussion

In this chapter, TALERS with various extents of N-terminal truncation of the TALE DNA-binding domain were tested for their ability to carry out binding, recombination and cleavage *in vivo* and *in vitro*. These TALERS were tested on substrates with different length spacers between the central 16 bp of the resolvase *res* site and the 17 bp DNA binding sites of the chosen TALE (T-20, T-22, T-24 and T-26). It was determined that the largest truncation of the N-terminal domain ( $\Delta 149$ ) was the most efficient of the TALERS tested, as only the  $\Delta 149$  protein was able to carry out cleavage *in vitro* and form what is hypothesised to be a synaptic tetramer-like structure in band-shift assays.

The failure of the longer TALER variants ( $\Delta 48$ ,  $\Delta 84$  and  $\Delta 119$ ) to form the larger protein complex and their inability to carry out cleavage assays may be due to the extra length at the N-terminal end of the TALE domain folding in a way that is detrimental to the ability of the protein to form the necessary protein-protein interactions for synapsis of the NM catalytic domains.

As there are no crystal structures available for TALE sequence N-terminal of the '-3' repeat, this region may be disordered, a state which may not be as detrimental to TALEN constructs, as the attachment site for the nuclease domain in such constructs is typically at the C-terminus of the TALE protein rather than the N-terminal attachment site utilised for these TALERS.

This C-terminal region is quite proline rich, and prolines are disruptors of secondary structures such as the  $\alpha$ -helices which form the repeat fingers in the TALE binding domains.

It is also possible that the inability of the TALER to be catalytically active may be due to there being an excess of amino acids between the DNA-binding domain and the catalytic domain, blocking the catalytic domain of NM resolvase from having access to the centre of the site in order to carry out catalysis. The  $\Delta 149$  TALER, with less protein between the catalytic domain and the DNA-binding domain is functional, and this may be because this construct places the catalytic domain in a position which permits access to the centre of the site for cleavage to take place.

It was also determined that of the T-site spacers tested, the shortest, (T-20) is not a viable site for the TALERS; the optimal spacing for the TALERS of the spacers tested were the T-22 and T-24 constructs, while the T-26 construct is less efficiently cleaved. The 22 base pair and 24 base pair spacing between the TALER binding sites was therefore shown to provide the correct distance to be accessible to the catalytic domain of the  $\Delta 149$  TALER. Interestingly, the optimal spacing observed for the TALERS is the same as the optimal spacing previously utilised for the ZFRs (Akopian et al 2003, Proudfoot et al 2011).

The three complexes seen for the  $\Delta 149$  construct are hypothesised to be monomer, dimer and tetramer complexes- a hypothesis which is supported by the evidence that the  $\Delta 149$  protein, the only protein seen to form this hypothesised tetramer complex, is also the only protein able to carry out cleavage. However, the presence of distinct monomer, dimer and tetramer complexes is not conclusively shown in this study. In order to confirm this hypothesis, assays similar to those carried out in chapter 3, utilising different lengths of labelled T-site, should be carried out to determine whether the putative tetramer complex does indeed contain two molecules of DNA.

Further work to characterise these TALER constructs would include the creation of more truncated variants, to determine whether further truncations could increase the activity of the protein beyond that seen in TALER  $\Delta 149$ , the shortest construct tested in this work and the addition of various flexible linker domains between the catalytic domain and the DNA binding domain in an attempt to optimise the positioning of the DNA binding domains relative to the catalytic domains when bound to the DNA. Additionally, the creation of T-sites with further different lengths of spacer between the DNA binding region and the centre of the site may provide a more optimal substrate for activity of TALERS, providing better spacing for the TALER to access the site and carry out recombination.

During the course of this study, another group reported TALER-mediated recombination activity, using gin invertase as a catalytic domain and the TALE DNA-binding domain AvrXa7 (Mercer et al 2012). The Gin-TALER work was carried out *in vivo*. They made both 3' and 5' attachments of the TALE AvrXa7 to Gin, demonstrating that, as hypothesised for the creation of my own TALER, only the attachment of TALE to the C-terminus of Gin was functional. They also carried out various N-terminal truncations, finding, as I did, that N-terminal truncation of the TALE DNA-binding domain produced more active variants. The recombination activity which they observed was not as good as that which their lab had previously observed with gin-ZFRs (Gaj et al 2010), with only approximately 7% recombination observed *in vivo*, even with the most active Gin-TALER variant.

So far TALERS have not proved to have the robust recombination activity observed with the ZFR proteins. The failure of the TALER constructs to recombine suggests that the positioning of the catalytic domain and DNA binding domains relative to the T-site is not yet optimal, preventing the necessary recombination complex from forming in a stable manner and thus preventing recombination. Hopefully further work will better characterise these proteins, leading to more efficient variants with activities analogous to the activity which has been observed to date with the ZFRs.

## Chapter 5: Conclusions and discussion

This work had two main goals; firstly, to examine the importance of the E-helix of Sin for synapsis at site I, and to determine the important residues in this interface, and secondly, to create a chimeric TALER construct, with the ability to be targeted to a site determined by the TALE DNA-binding domain.

Chapter 3 discussed the sequence selectivity of Sin resolvase-mediated binding and synapsis of site I. It was shown that altering the nucleotides at the centre of site I from the Sin nucleotides to the corresponding Tn3 nucleotides negatively effects binding by Sin; highlighting the importance of recognition of these bases by the protein. It was also demonstrated that both halves of the site are required for the formation of dimer and synaptic tetramer complexes of resolvase.

Assays with the truncated ‘tadpole’ variants show that those with the E-helix intact are competent for binding and synapsis, but deletion of the residues between 109 and 124 of the E-helix blocks the formation of a synapse, highlighting the importance of this region of the protein for formation of a tetramer of E-helices.

Swapping the ‘arm’ region of Sin with the equivalent residues from Tn3 resolvase prevented the formation of a synaptic complex on the Sin site I, again demonstrating that residues within the E-helix ‘arm’ region do have specificity for binding at the site. The ability of the longest ‘tadpole’ ( $\Delta 101$ ) to form a synapse at the site was unexpected, and may be due to co-operative binding and the E-helix interactions of the protein conferring stability to the protein. Blocking of the formation of the synapse, both by alteration of the ‘arm’ region of the E-helix and by alteration of the DNA sequence at the centre of the site, demonstrates that the sequence at the centre of the *res* site

is important, and that the 'arm' region of resolvase is making specific contacts with these nucleotides.

Mutations made within the E-helix, (between residues 110 and 121) both on the flat, hydrophobic interface between the E-helices and on the outside of the interface (as discussed in section 3.6) demonstrated that, as predicted from the crystal structure data, residues on this interface are important to synapsis. From the crystal structures, it can be additionally seen that these E-helix residues have potential interactions between the E-helices and the catalytic domains, either in cis (I114 and S121) or in trans (L112, I116 and M119) and between the anti-parallel E-helices (L112, I116 and M119), forming interactions which hold the 'rotational dimers' together. These mutations also highlighted the similarity in behaviour between these 'tadpole' constructs and their full-length activated counterparts, lending credence to the hypothesis that the E-helix bundle at the centre of the tetramer, discussed in chapter 3, acts as a basic synapsis-competent module.

Chapter 4 concerned the creation of chimeric TALER proteins, and demonstrated that it is possible to create a TALER with the ability to bind to its target site and which is able to carry out cleavage *in vitro*. However, the Tn3-TALER created in this study was not observed to have recombination activity. Further optimization of the linker length and site spacing may yield more efficient TALER variants with activities analogous to those which have been observed to date with the ZFRs.

The TALER constructs discussed in this work used a single TALE DNA-binding domain, targeted to identical sites flanking the Tn3 resolvase *res* site. In order for TALER constructs to be targeted to a desired region within genomic sequence, creation of heterodimeric TALERs, which are able to target different DNA-sites would be desirable. Such heterodimeric constructs would be useful for targeting biologically relevant sites within the genome.



The work discussed in chapter 3 has highlighted the sequence selectivity of the resolvase 'arm' region, and it is also known that the catalytic domain of resolvase has sequence selectivity for the centre of the site. A further goal for the creation of chimeric recombinases such as TALERs for use as tools in biotechnology or genetic surgery would be the creation of chimeric proteins with the selectivity of the recombinase domain removed, creating a non-selective recombinase, which is fully reliant on the sequence specificity inherent to the DNA-binding domain of the TALE. Such a protein would provide an alternative to the TALENs which have been widely discussed in the literature. This approach however would require a vast amount of complex protein engineering, may prove very difficult, and is a goal for the future. Future work to further examine the important residues in the catalytic module and the E-helix of resolvase may shed more light on ways to make more efficient chimeric TALE recombinases with potential applications in biotechnology and genetic engineering.

## References:

- Akopian, A., He, J., Boocock, M. R., & Stark, W. M. (2003): 'Chimeric recombinases with designed DNA sequence recognition.' *Proceedings of the National Academy of Sciences*, 100(15), 8688-8691.
- Akopian, A., & Marshall Stark, W. (2005): 'Site-specific DNA recombinases as instruments for genomic surgery'. *Advances in Genetics*, 55, 1-23.
- Arnold, P. H., Blake, D. G., Grindley, N. D., Boocock, M. R., & Stark, W. M. (1999): 'Mutants of Tn3 resolvase which do not require accessory binding sites for recombination activity.' *The EMBO Journal*, 18(5), 1407-1414.
- Ariyoshi, M., Nishino, T., Iwasaki, H., Shinagawa, H. and Morikawa, K. (2000): 'Crystal structure of the Holliday junction DNA in complex with a single RuvA tetramer', *Proceedings of the National Academy of Sciences*, 97(15), 8257-62.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S. & Horvath, P. (2007): 'CRISPR provides acquired resistance against viruses in prokaryotes.' *Science*, 315(5819), 1709-1712.
- Beerli, R. R., & Barbas, C. F. (2002): 'Engineering polydactyl zinc-finger transcription factors.' *Nature Biotechnology*, 20(2), 135-141.
- Blake, D.G. (1993): 'Binding of Tn3 resolvase to its recombination site.' PhD thesis University of Glasgow.
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S. & Bonas, U. (2009): 'Breaking the code of DNA binding specificity of TAL-type III effectors.' *Science*, 326(5959), 1509-1512.
- Briggs, A. W., Rios, X., Chari, R., Yang, L., Zhang, F., Mali, P., & Church, G. M. (2012): 'Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers.' *Nucleic Acids Research*.
- Burke, M. E., Arnold, P. H., He, J., Wenwieser, S. V., Rowland, S. J., Boocock, M. R. and Stark, W. M. (2004): 'Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation.', *Molecular Microbiology*, 51(4), 937-48.

Camerini-Otero, R. D. and Hsieh, P. (1995): 'Homologous recombination proteins in prokaryotes and eukaryotes.', *Annual Review of Genetics*, 29, 509-52.

Cavazzana-Calvo, M., Payen, E., Negre, O., Wang, G., Hehir, K., Fusil, F. & Leboulch, P. (2010): 'Transfusion independence and hmga2 activation after gene therapy of human [bgr]-thalassaemia.' *Nature*, 467(7313), 318-322.

Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C. & Voytas, D. F. (2011): 'Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting.' *Nucleic Acids Research*, Vol. 39.

Chambers, S. P., Prior, S. E., Earstow, D. A. and Minton, N. P. (1988): 'pMTL nick-cloning vectors. I. improved pUC polylinker regions to facilitate the use of Sonicated DNA for nucleotide sequencing. *Gene* 68, 139-149.

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A. & Voytas, D. F. (2010): 'Targeting DNA double-strand breaks with TAL effector nucleases.' *Genetics*, 186(2), 757-761.

Collier, L. S., Carlson, C. M., Ravimohan, S., Dupuy, A. J., & Largaespada, D. A. (2005): 'Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse.' *Nature*, 436(7048), 272-276.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N. & Zhang, F. (2013): 'Multiplex genome engineering using CRISPR/Cas systems.' *Science*, 339(6121), 819-823.

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J. K. & Yan, N. (2012): 'Structural basis for sequence-specific recognition of DNA by TAL effectors.' *Science*, 335(6069), 720-723.

Dreier, B., Beerli, R. R., Segal, D. J., Flippin, J. D., & Barbas, C. F. (2001): 'Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors.' *Journal of Biological Chemistry*, 276(31), 29466-29478.

Dreier, B., Fuller, R. P., Segal, D. J., Lund, C. V., Blancafort, P., Huber, A., ... & Barbas, C. F. (2005): 'Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors.' *Journal of Biological Chemistry*, 280(42).

Fischer, A., Hacein-Bey-Abina, S. & Cavazzana-Calvo, M. (2011): 'Gene therapy for primary adaptive immune deficiencies.' *Journal of Allergy and Clinical Immunology*, 127(6), 1356-1359.

Gaj, T., Guo, J., Kato, Y., Sirk, S. J., & Barbas III, C. F. (2012): 'Targeted gene knockout by direct delivery of zinc-finger nuclease proteins.' *Nature Methods*, 9(8), 805-807.

Gaj, T., Gersbach, C. A., & Barbas III, C. F. (2013): 'ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering.' *Trends in Biotechnology*, 31(7), 397-405.

Gersbach, C. A., Gaj, T., Gordley, R. M., Mercer, A. C., & Barbas, C. F. (2011): 'Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase.' *Nucleic Acids Research*, 39(17), 7868-7878.

Gordley, R. M., Smith, J. D., Gräslund, T. & Barbas III, C. F. (2007): 'Evolution of programmable zinc finger-recombinases with activity in human cells.' *Journal of Molecular Biology*, 367(3), 802-813.

Gordley, R. M., Gersbach, C. A., & Barbas, C. F. (2009): 'Synthesis of programmable integrases.' *Proceedings of the National Academy of Sciences*, 106(13), 5053-5058.

Grindley, N.D.F., (1994): 'Resolvase-mediated site-specific recombination'. *Nucleic Acids Molecular Biology* 8, pp. 236–267.

Grindley N. D. F., Craig N.L, Craigie R, Geller M, Lambowitz AM (2002): 'The movement of Tn3-like elements: 'Transposition and cointegrate resolution'. *Mobile DNA II. Washington (D. C.): American Society for Microbiology Press. pp. 272–302.*

Grindley, N. D., Whiteson, K. L. and Rice, P. A. (2006): 'Mechanisms of site-specific recombination.', *Annual Review of Biochemistry*, 75, 567-605.

Guo, J., Gaj, T., & Barbas III, C. F. (2010): 'Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases.' *Journal of molecular biology*, 400(1), 96-107.

Holliday, R. (1964): 'A mechanism for gene conversion in fungi.', *Genetic Research*, 89(5-6), 285-307.

- Ivics, Z., Katzer, A., Stüwe, E. E., Fiedler, D., Knespel, S., & Izsvák, Z. (2007): 'Targeted Sleeping Beauty transposition in human cells.' *Molecular Therapy*, 15(6), 1137-1144.
- Keenholtz, R. A., Rowland, S. J., Boocock, M. R., Stark, W. M. and Rice, P. A. (2011) 'Structural basis for catalytic activation of a serine recombinase.', *Structure*, 19(6), 799-809.
- Kim, Y., Kweon, J., Kim, A., Chon, J. K., Yoo, J. Y., Kim, H. J. & Kim, J. S. (2013): 'A library of TAL effector nucleases spanning the human genome.' *Nature Biotechnology*, 31(3), 251-258.
- Kolb, A. F., Coates, C. J., Kaminski, J. M., Summers, J. B., Miller, A. D., & Segal, D. J. (2005): 'Site-directed genome modification: nucleic acid and protein modules for targeted integration and gene correction.' *Trends in Biotechnology*, 23(8), 399-406.
- Laemmli, U. K. (1970): 'Cleavage of structural proteins during the assembly of the head of bacteriophage T4.' *Nature* 227, 680-685
- Li, W., Kamtekar, S., Xiong, Y., Sarkis, G. J., Grindley, N. D. and Steitz, T. A. (2005): 'Structure of a synaptic gammadelta resolvase tetramer covalently linked to two cleaved DNAs.', *Science*, 309(5738).
- Li, T., Huang, S., Jiang, W. Z., Wright, D., Spalding, M. H., Weeks, D. P., & Yang, B. (2011): 'TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain.' *Nucleic Acids Research*, 39(1), 359-372.
- Maeder, M. L., Thibodeau-Beganny, S., Osiak, A., Wright, D. A., Anthony, R. M., Eichinger, M. & Joung, J. K. (2008): 'Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification.' *Molecular Cell*, 31(2), 294-301.
- Mak, A. N. S., Bradley, P., Cernadas, R. A., Bogdanove, A. J., & Stoddard, B. L. (2012): 'The crystal structure of TAL effector PthXo1 bound to its DNA target.' *Science*, 335(6069), 716-719.
- Mercer, A. C., Gaj, T., Fuller, R. P., & Barbas, C. F. (2012): 'Chimeric TALE recombinases with programmable DNA sequence specificity.' *Nucleic Acids Research*, 40(21), 11163-11172.

- Mercer, A. C., Gaj, T., Sirk, S. J., Lamb, B. M., & Barbas III, C. F. (2013): 'Regulation of endogenous human gene expression by ligand-inducible TALE transcription factors.' *ACS Synthetic Biology*.
- Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y. L., Rupniewski, I. & Rebar, E. J. (2007): 'An improved zinc-finger nuclease architecture for highly specific genome editing.' *Nature Biotechnology*, 25(7), 778-785.
- Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., ... & Rebar, E. J. (2011): 'A TALE nuclease architecture for efficient genome editing.' *Nature Biotechnology*, 29(2), 143-148.
- Moehle, E. A., Rock, J. M., Lee, Y. L., Jouvenot, Y., DeKolver, R. C., Gregory, P. D. & Holmes, M. C. (2007): 'Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases.' *Proceedings of the National Academy of Sciences*, 104(9), 3055-3060.
- Morrison, K. L., Weiss, G. A. (2001): "Combinatorial alanine-scanning". *Current Opinion in Chemical Biology* 5 (3): 302–307
- Moscou, M. J., & Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science*, 326(5959), 1501-1501.
- Mouw, K. W., Rowland, S. J., Gajjar, M. M., Boocock, M. R., Stark, W. M. and Rice, P. A. (2008): 'Architecture of a serine recombinase-DNA regulatory complex.', *Molecular Cell*, 30(2), 145-155.
- Mouw, K. W., Steiner, A. M., Ghirlando, R., Li, N. S., Rowland, S. J., Boocock, M. R., Stark, W. M., Piccirilli, J. A. and Rice, P. A. (2010): 'Sin resolvase catalytic activity and oligomerization state are tightly coupled.', *Journal of Molecular Biology*, 404(1), 16-33.
- Mussolino, C., & Cathomen, T. (2012): 'TALE nucleases: tailored genome engineering made easy.' *Current Opinion in Biotechnology*, 23(5), 644-650.
- Nöllmann, M., He, J., Byron, O. and Stark, W., *Solution structure of the Tn3 resolvase-crossover site synaptic complex. Molecular Cell*, 2004, 16 (1), 127-137.
- Olorunniji, F. J., He, J., Wenwieser, S. V., Boocock, M. R. and Stark, W. M. (2008): 'Synapsis and catalysis by activated Tn3 resolvase mutants.', *Nucleic Acids Research*, 36(22), 7181-7191.

- Olorunniji, F. J. and Stark, W. M. (2009): 'The catalytic residues of Tn3 resolvase.', *Nucleic Acids Research*, 37(22), 7590-7602.
- Olorunniji, F. J. and Stark, W. M. (2010): 'Catalysis of site-specific recombination by Tn3 resolvase.', *Biochemical Society Transactions*, 38(2), 417-421.
- Ott, M. G., Schmidt, M., Schwarzwaelder, K., Stein, S., Siler, U., Koehl, U. & Grez, M. (2006): 'Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1.' *Nature Medicine*, 12(4), 401-409.
- Paulsen, I. T., Gillespie, M. T., Littlejohn, T. G., Hanvivatvong, O., Rowland, S. J., Dyke, K. G. and Skurray, R. A. (1994): 'Characterisation of *sin*, a potential recombinase-encoding gene from *Staphylococcus aureus*.' *Gene*, 141(1), 109-114.
- Prorocic, M. M., Wenlong, D., Olorunniji, F. J., Akopian, A., Schloetel, J. G., Hannigan, A. & Stark, W. M. (2011): 'Zinc-finger recombinase activities in vitro.' *Nucleic Acids Research*, gkr652.
- Proudfoot, C., McPherson, A. L., Kolb, A. F., & Stark, W. M. (2011): 'Zinc finger recombinases with adaptable DNA sequence specificity.' *PLoS one*, 6(4), e19537.
- Rad, R., Rad, L., Wang, W., Cadinanos, J., Vassiliou, G., Rice, S. & Bradley, A. (2010): 'PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice.' *Science*, 330(6007), 1104-1107.
- Reyon, D., Tsai, S. Q., Khayter, C., Foden, J. A., Sander, J. D., & Joung, J. K. (2012): 'FLASH assembly of TALENs for high-throughput genome editing.' *Nature Biotechnology*, 30(5), 460-465.
- Rice, P. A., & Steitz, T. A. (1994): 'Model for a DNA-mediated synaptic complex suggested by crystal packing of gamma delta resolvase subunits.' *The EMBO Journal*, 13(7), 1514-1524.
- Rice, P. A., Mouw, K. W., Montañó, S. P., Boocock, M. R., Rowland, S. J. and Stark, W. M. (2010): 'Orchestrating serine resolvases.', *Biochemical Society Transactions*, 38(2), 384-387.

- Rowland, S. J., Stark, W. M. and Boocock, M. R. (2002): 'Sin recombinase from *Staphylococcus aureus*: synaptic complex architecture and transposon targeting.', *Molecular Microbiology*, 44(3), 607-619.
- Rowland, S. J., Boocock, M. R. and Stark, W. M. (2005): 'Regulation of Sin recombinase by accessory proteins.', *Molecular Microbiology*, 56(2), 371-382.
- Rowland, S. J., Boocock, M. R. and Stark, W. M. (2006): 'DNA bending in the Sin recombination synapse: functional replacement of HU by IHF', *Molecular Microbiology*, 59(6), 1730-1743.
- Rowland, S. J., Boocock, M. R., McPherson, A. L., Mouw, K. W., Rice, P. A. and Stark, W. M. (2009): 'Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome.', *Molecular Microbiology*, 74(2), 282-298.
- Samson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y. L., & Weiss, D. S. (2013): 'A CRISPR/Cas system mediates bacterial innate immune evasion and virulence.' *Nature*, 497(7448), 254-257.
- Sander, J. D., Cade, L., Khayter, C., Reyon, D., Peterson, R. T., Joung, J. K., & Yeh, J. R. J. (2011): 'Targeted gene disruption in somatic zebrafish cells using engineered TALENs'. *Nature Biotechnology*, 29(8), 697-698.
- San Filippo, J. Sung, P. and Klein. H (2008): 'Mechanism of eukaryotic homologous recombination': *Annual Review of Biochemistry Vol. 77*: 229-257
- Sarkis, G. J., Murley, L. L., Leschziner, A. E., Boocock, M. R., Stark, W. M. and Grindley, N. D. (2001): 'A model for the gamma delta resolvase synaptic complex.', *Molecular Cell*, 8(3), 623-631.
- Schmid-Burgk, J. L., Schmidt, T., Kaiser, V., Höning, K., & Hornung, V. (2013): 'A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes.' *Nature Biotechnology*, 31(1), 76-81.
- Schneider, F., Schwikardi, M., Muskhelishvili, G., & Dröge, P. (2000): 'A DNA-binding domain swap converts the invertase Gin into a resolvase.' *Journal of Molecular Biology*, 295(4), 767-775.



- Sirk, S. J., Gaj, T., Jonsson, A., Mercer, A. C., & Barbas, C. F. (2014): 'Expanding the zinc-finger recombinase repertoire: directed evolution and mutational analysis of serine recombinase specificity determinants.' *Nucleic Acids Research*, 42(7), 4755-4766.
- Smith, M. C. and Thorpe, H. M. (2002) 'Diversity in the serine recombinases.', *Molecular Microbiology*, 44(2), 299-307.
- Stark, W. M., Boocock, M. R. and Sherratt, D. J. (1992): 'Catalysis by site-specific recombinases.', *Trends Genet*, 8(12), 432-439.
- Stark, W. M. and Boocock, M. R. (1995): 'Gatecrashers at the catalytic party.', *Trends in Genetics*, 11(4), 121-123.
- Streubel, J., Blücher, C., Landgraf, A., & Boch, J. (2012): 'TAL effector RVD specificities and efficiencies.' *Nature Biotechnology*, 30(7), 593-595.
- Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990): 'Use of T7 RNA polymerase to direct expression of cloned genes.' *Methods in Enzymology*, 185, 60-89.
- Summers, D. K., & Sherratt, D. J. (1988): 'Resolution of ColE1 dimers requires a DNA sequence implicated in the three-dimensional organization of the *cer* site.' *The EMBO Journal*, 7(3), 851-856.
- Sun, N., Liang, J., Abil, Z., & Zhao, H. (2012): 'Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease.' *Molecular BioSystems*, 8(4), 1255-1263.
- Thyagarajan, B., Liu, Y., Shin, S., Lakshmipathy, U., Scheyhing, K., Xue, H. & Chesnut, J. D. (2008): 'Creation of engineered human embryonic stem cell lines using phiC31 integrase.' *Stem Cells*, 26(1), 119-126.
- Tsuji, S., Futaki, S., & Imanishi, M. (2013): 'Creating a TALE protein with unbiased 5'-T binding.' *Biochemical and Biophysical Research Communications*, 441(1), 262-265.
- Urnov, F. D., Miller, J. C., Lee, Y. L., Beausejour, C. M., Rock, J. M., Augustus, S. & Holmes, M. C. (2005): 'Highly efficient endogenous human gene correction using designed zinc-finger nucleases.' *Nature*, 435(7042), 646-651.

Wirth, D., Gama-Norton, L., Riemer, P., Sandhu, U., Schucht, R., & Hauser, H. (2007): 'Road to precision: recombinase-based targeting technologies for genome engineering.' *Current Opinion in Biotechnology*, 18(5), 411-419.

Yang, W. and Steitz, T. A. (1995): 'Crystal structure of the site-specific recombinase gamma delta resolvase complexed with a 34 bp cleavage site', *Cell*, 82(2), 193-207.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G. M., & Arlotta, P. (2011): 'Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription.' *Nature Biotechnology*, 29(2), 149-153.

```

SpeI                                     PfoI
|                                     48 |
1
ACTAGTGATCCCATTCTGTTGCGGCACGCCAAGTCCTGCCGCGAGCTTCTGCCCGGACCCCAACCGGATAGTGTACAGCCGACTGCAGATCGGGGGGGGGCTCCGCCTGCTGGCGGCCCCCTGGATGGCTTGCCCCTCGGCGGACGATGTCCTGG
T S D P I R S R T P S P A R E L L P G P Q P D S V Q P T A D R G G A P P A G G P L D G L P A R R T M S R

                                     84                                     SacII
                                     |
ACCCGGCTGCCATCTCCCCCTGCGCCCTCGCCTGCGTTCTCGGCGGGCAGCTTCAGCGATCTGCTCCGTCA GTTCGATCCGTCGCTTCTTGATACATCGCTTCTTGATTCGATGCCCCGCGGTCGGCACGCCGCATACAGCGGCTGCCCCAGCAGAG
T R L P S P P A P S P A F S A G S F S D L L R Q F D P S L L D T S L L D S M P A V G T P H T A A A P A E

                                     BsiWI                                     SalI
                                     119                                     149
                                     |                                     |
TGC GATGAGGTGCAATCGGGTCTGCGTGCAGCCGATGACCCGCCACCCACCGTACGTGTGCTGTCACTGCGGCGCGACCGCCGCGGCCAAGCCGGCCCCGCGACGGCGTGC GGGCGCAACCCTCCGACGCTTCGCCGGCAGCGCAGGTCGACTTG
C D E V Q S G L R A A D D P P P T V R V A V T A A R P P R A K P A P R R R A A Q P S D A S P A A Q V D L

AGGACCCTAGGTTATTGCAACAGCAACAGGAGAAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTGCGCTTTCACAGCACCTGCGGCGCTTGGGACGGTGGCT
R T L G Y S Q Q Q Q E K I K P K V R S T V A Q H H E A L V G H G F T H A H I V A L S Q H P A A L G T V A

GTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCGATTGTAGGGGTGCGTAAACAGTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAG
V K Y Q D M I A A L P E A T H E A I V G V G K Q W S G A R A L E A L L T V A G E L R G P P L Q L D T G Q

                                     ApaLI
                                     |
CTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCAC
L L K I A K R G G V T A V E A V H

```

## Supplementary 1: GeneArt construct TAL\_SpeI-ApaLI

The full nucleotide sequence of the synthetic construct TAL\_SpeI-ApaLI ordered from MWG for construction of the TALER expression plasmid. The location of the introduced silent restriction sites and the truncation points which they introduce is shown.

```

MfeI
|
CAATTGTCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGAATTCCCGAG
Q L S R P D P A L A A L T N D H L V A L A C L G G R P A L D A V K K G L P H A P A L I K R T N R R I P E

                                     KpnI
                                     |
AGAACTTCCCACCGAGTGGCGCATCACCATCACCATCACTGATGATCAGGTACC
R T S H R V A H H H H H H * * S G T

```

## Supplementary 2: GeneArt construct TAL\_MfeI-KpnI

The full nucleotide sequence of the synthetic construct TAL\_MfeI-KpnI ordered from MWG for construction of the TALER expression plasmid. The location of the introduced histidine tag and termination codons is shown.